DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Two-Dimensional Plasmonics in Massive and Massless Electron Gases

Citation
Yoon, Hosang. 2014. Two-Dimensional Plasmonics in Massive and Massless Electron Gases. Doctoral dissertation, Harvard University.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:13070026

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Two-Dimensional Plasmonics in Massive and Massless Electron Gases

A dissertation presented

by

Hosang Yoon

to

The School of Engineering and Applied Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Applied Physics

Harvard University
Cambridge, Massachusetts

June 2014

©2014 Hosang Yoon

All rights reserved.

Thesis advisor

Donhee Ham

Author

Hosang Yoon

Two-Dimensional Plasmonics in Massive and Massless Electron Gases

Abstract

Plasmonic waves in solid-state are caused by collective oscillation of mobile charges inside or at the surface of conductors. In particular, surface plasmonic waves propagating at the skin of metals have recently attracted interest, as they reduce the wavelength of electromagnetic waves coupled to them by up to $\sim$10 times, allowing one to create miniaturized wave devices at optical frequencies. In contrast, plasmonic waves on two-dimensional (2D) conductors appear at much lower infrared and THz-GHz frequencies, near or in the electronics regime, and can achieve far stronger wavelength reduction factor reaching well above 100. In this thesis, we study the unique machinery of 2D plasmonic waves behind this ultra-subwavelength confinement and explore how it can be used to create various interesting devices.

To this end, we first develop a physically intuitive theoretical formulation of 2D plasmonic waves, whose two main components—the Coulomb restoration force and inertia of the collectively oscillating charges—are combined into a transmission-line-like model. We then use this formulation to create various ultra-subwavelength 2D plasmonic devices. For the 2D conductor, we first choose GaAs/AlGaAs heterostructure—a 2D electron gas consisting of massive ($m^* > 0$) electrons—demonstrating plasmonic bandgap crystals, interferometers, and negatively refracting metamaterials. We then examine a 2D plasmonic device based on graphene, a 2D electron gas consisting of effectively massless ($m^* = 0$) electrons. We theoretically show and experimentally demonstrate that the massless electrons in graphene can surprisingly exhibit a collective mass when subjected to a collective excitation, providing the inertia that is essential for the propagation of 2D plasmonic waves.

Lastly, we theoretically investigate the thermal current fluctuation behaviors in massive and massless electron gases. While seemingly unrelated on first sight, we show that the thermal current fluctuation is

actually intimately linked to the collective mass of the massive or massless electron gas. Thus, we show that the thermal current fluctuation behaviors can also be described by the same theoretical framework introduced earlier, suggesting a possibility to design new concept devices and experiments based on this linkage.

Contents

List of Figures

List of Tables

Citations to Previously Published Work

Large portions of this thesis have appeared in the following papers in publication:

H. Yoon, K. Y. M. Yeung, P. Kim, and D. Ham, "Plasmonics with two-dimensional conductors," *Philosophical Transactions of the Royal Society A*, vol. 372, p. 20130104, 2014.

W. F. Andress, H. Yoon, K. Y. M. Yeung, L. Qin, K. West, L. Pfeiffer, and D. Ham, "Ultra-subwavelength two-dimensional plasmonic circuits," *Nano Letters*, vol. 12, pp. 2272-2277, 2012.

H. Yoon, K. Y. M. Yeung, V. Umansky, and D. Ham, "A Newtonian approach to extraordinarily strong negative refraction," *Nature*, vol. 488, pp. 65-69, 2012.

H. Yoon, C. Forsythe, L. Wang, N. Tombros, K. Watanabe, T. Taniguchi, J. Hone, P. Kim, and D. Ham, "Measurement of collective dynamical mass of Dirac fermions in graphene," *Nature Nanotechnology*, 2014 (doi:10.1038/nnano.2014.112).

H. Yoon and D. Ham, "Massive thermal fluctuation of massless graphene electrons," *arXiv:1405.2356 [cond-mat]* 2014.

Acknowledgments

First and foremost, I would like to thank my advisor, Donhee Ham, for his tremendous support and constant encouragement of my research and career development.

I would also like to thank my group members: William Andress for his guidance in experiments and fabrication; Kitty Yeung and Ling Qin for their collaborations and fruitful discussions; Dongwan Ha and Guangyu Xu for their advice and assistance.

In addition, I would like to give special thanks to my colleagues: Philip Kim and his group members, Nikolaos Tombros and Carlos Forsythe, for providing generous support for my research on graphene plasmonic devices; Kenji Watanabe and Takashi Taniguchi for providing high quality hexagonal boron nitride samples; Loren Pfeiffer and Vladimir Umansky for providing excellent samples for my research on plasmonic devices with two-dimensional electron gas.

Finally, I would like to thank professors Marko Lončar, Hongkun Park, Federico Capasso, Robert Westervelt, and Roger Brockett for their guidance and support throughout my Ph.D. study.

Chapter 1

Introduction

In this thesis, we will examine the unique characteristics of two-dimensional (2D) plasmonic waves and explore the various device applications enabled by them. Before going into the details, we will first provide an introduction of the 2D electronic systems to be discussed in this thesis, and then briefly discuss and compare the plasmonic waves that appear in various dimensions to motivate the main theme of this thesis.

1.1 2D Electronic Systems

A 2D electronic system is defined as a collection of electrons (or holes) that can move freely in two dimensions, which we denote x and y, while their movement is restricted in the third dimension, which we denote z. These 2D electronic systems are found in a number of different structures such as the inversion layer of MOSFET, specifically engineered semiconductor heterojunctions, surface of liquid helium, and graphene. In this thesis, we are interested in two of these systems, namely, semiconductor heterojunction based 2D electron gas (2DEG) and graphene, as model 2D electronic systems consisting of massive and effectively massless electrons.

A semiconductor heterojunction 2DEG is formed by utilizing the sudden change in the conduction band energy level in the vicinity of a junction between two different semiconductor materials due to the mismatch of bandgaps of these materials. The layer structure and band diagram of a typical GaAs/AlGaAs 2DEG, which is the 2DEG to be used in this thesis, is shown in Figs. 1.1a,b [1]. By

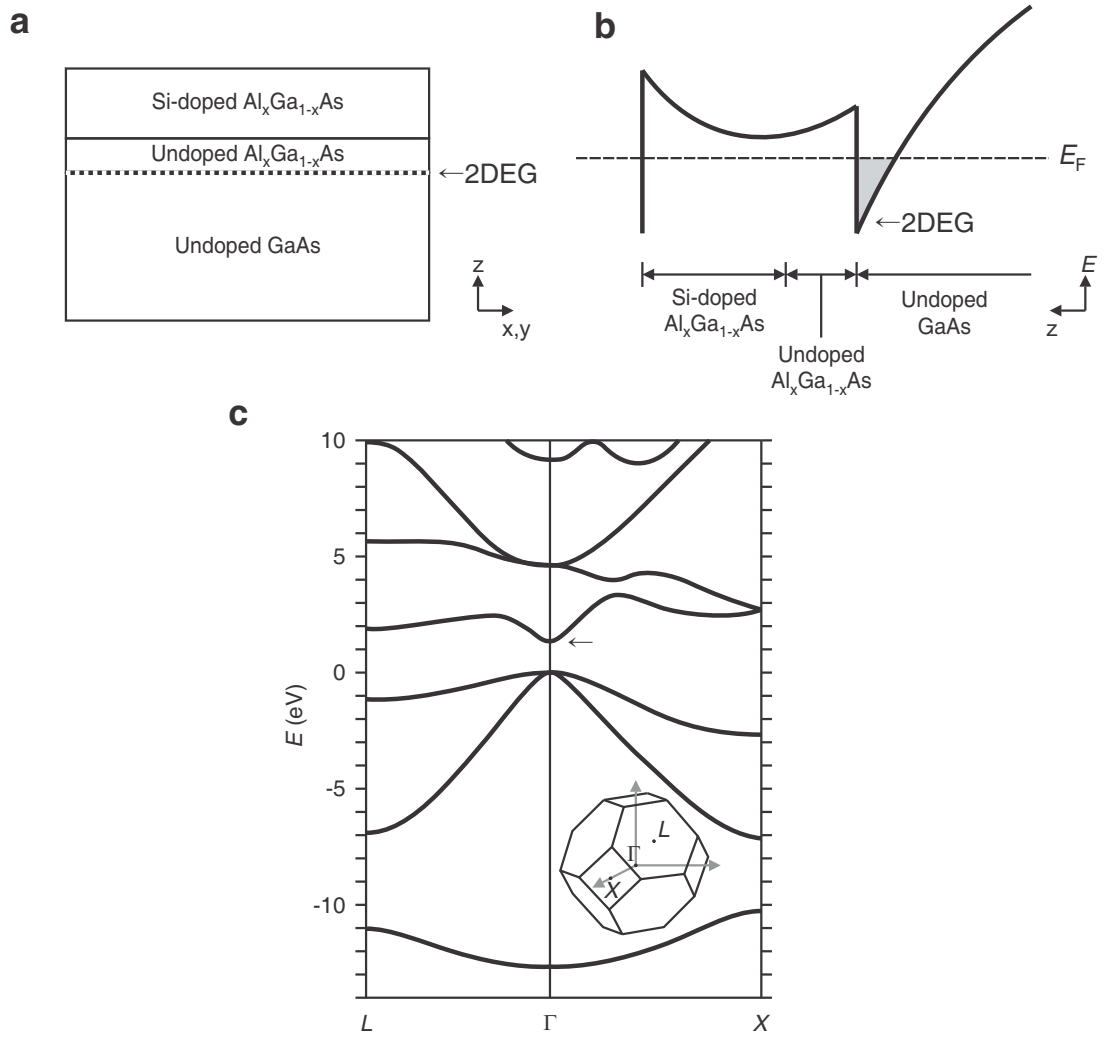

Figure 1.1: **a**, Typical vertical structure of GaAs/AlGaAs 2DEG. **b**, Band diagram of the structure in **a** [1]. Only the conduction band is shown. **c**, Band structure of GaAs (inset: Brillouin zone in the momentum space) [2].

carefully designing the structure and tuning the doping of the top layer, the bands can be engineered so that only the lowest energy level in z direction is occupied in the 2D well at the junction (shaded gray in Fig. 1.1b), thereby forming a 2D electronic system in x and y directions. This well, typically of ~ 10 nm thickness, resides in the GaAs side. Within the well, the conduction in x and y directions is described by an approximate form of the lowest part of the conduction band of GaAs near Γ point (Fig. 1.1c, arrow [2]), given by

$$E(\vec{k}) = \frac{\hbar^2 k^2}{2m^*}, \tag{1.1}$$

where $k \equiv |\vec{k}| = k_x^2 + k_y^2$ and $m^* \approx 0.067 m_{\mathrm{e}}$ for GaAs (m_{e}: electron mass) [1]. Here, $\vec{k}$ is relative to Γ point, and $E(\vec{k})$ is relative to the energy at Γ point. This description then becomes identical to that of a collection of electrons in 2D with an effective mass m^*.

Graphene is a single[1] layer of carbon atoms connected to each other in a honeycomb lattice (Fig. 1.2a). Unlike the semiconductor heterojunction 2DEG, which is a 2D well formed in a three-dimensional (3D) structure by carefully engineering the band in z direction, graphene is naturally a 2D electronic system because of its physical structure being a single-atom-thick 2D conductor. The band structure calculated in the tight-binding approach for the structure in Fig. 1.2a [3] reveals that the conduction band exhibits minimum at two inequivalent points in the Brillouin zone, K and K' (also known as Dirac points), and the valence band shares the same shape with an inverted sign (Figs. 1.2b,c). More interestingly, the conduction and valence bands approach Dirac points with a linear slope, leading to an approximate form of the bands near Dirac points given by

$$E(\vec{k}) = \pm \hbar v_{\mathrm{F}} k, \tag{1.2}$$

where $v_{\mathrm{F}} \approx 10^6$ m/s is a constant proportional to the slope of the band structure near Dirac points. Here, $\vec{k}$ is relative to a Dirac point (K or K'), and $E(\vec{k})$ is relative to the energy at Dirac points. At low enough energies, the bands near K and K' points become degenerate, which can be accounted for in calculations as a valley degeneracy of 2.

The behaviors of the low energy expressions for the band structures of GaAs/AlGaAs 2DEG and graphene, which we refer to as E-k dispersion relation, are compared in Fig. 1.3. We can see that, as opposed to the picture of electrons behaving as massive particles with a quadratic E-k relation in

[1] We only consider monolayer graphene that exhibits zero effective electron mass in this thesis.

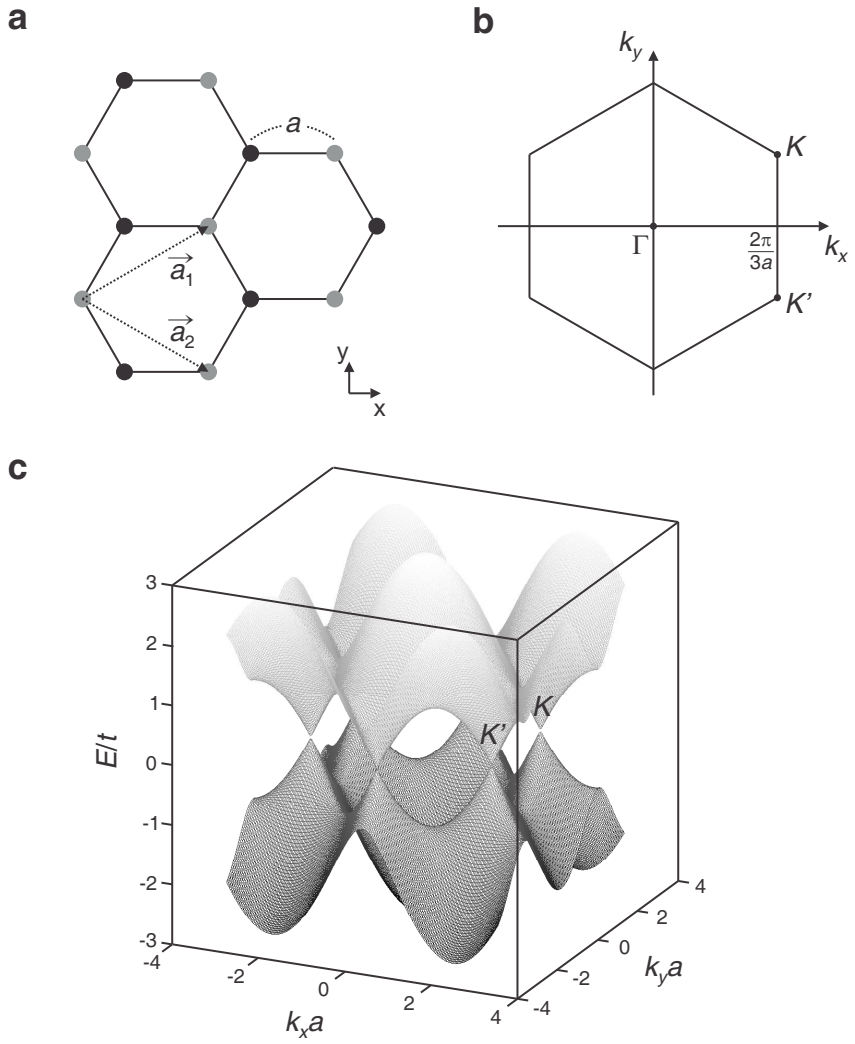

Figure 1.2: **a**, Honeycomb lattice structure of monolayer graphene. There are two atoms, colored black and gray, in each unit cell. **b**, Brillouin zone in the momentum space. **c**, Band structure calculated in the tight-binding approximation [3]. $a \approx 0.142$ nm is the carbon-carbon distance and $t \approx 2.5$ eV is the nearest-neighbor hopping amplitude in the tight-binding approximation.

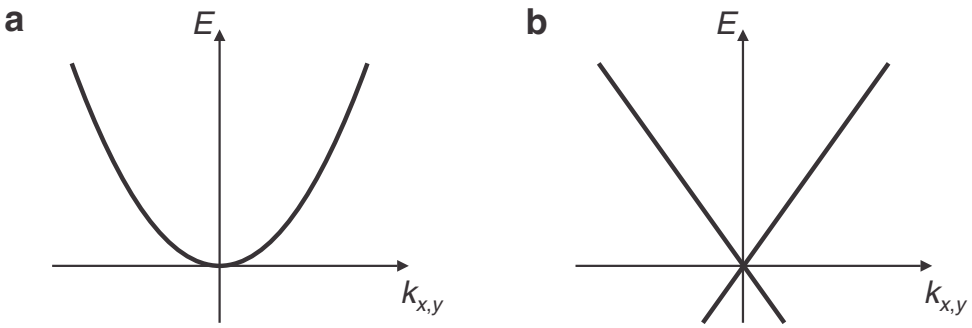

Figure 1.3: a, Simplified 2D band structure of GaAs/AlGaAs 2DEG near Γ point. **b**, Simplified 2D band structure of monolayer graphene near a Dirac point (K or K').

Table 1.1: E_{F}, k_{F}, v_{F}, and $D(E)$ for 2DEG, graphene, and 3D metal (free electron model) at $T = 0$ K. g is the spin and valley degeneracy ($g = 2$ for GaAs/AlGaAs 2DEG; $g = 4$ for graphene; $g = 2$ for 3D metal).

	E_{F}	k_{F}	v_{F}	$D(E)$
2DEG	$\frac{\hbar^2 k_{\mathrm{F}}^2}{2m^*}$	$\left(\frac{4\pi n_0}{g}\right)^{1/2}$	$\frac{\hbar k_{\mathrm{F}}}{m^*}$	$\frac{gm^*}{2\pi\hbar^2}$
Graphene	$\hbar v_{\mathrm{F}} k_{\mathrm{F}}$	$\left(\frac{4\pi n_0}{g}\right)^{1/2}$	v_{F}	$\frac{gE}{2\pi(\hbar v_{\mathrm{F}})^2}$
3D metal	$\frac{\hbar^2 k_{\mathrm{F}}^2}{2m^*}$	$\left(\frac{6\pi^2 n_{3\mathrm{D}}}{g}\right)^{1/3}$	$\frac{\hbar k_{\mathrm{F}}}{m^*}$	$\frac{gm^*}{2\pi^2\hbar^2}\sqrt{\frac{2m^*E}{\hbar^2}}$

GaAs/AlGaAs 2DEG, the graphene electrons behave effectively as massless particles with a linear E-k relation. In addition, while the Γ point of GaAs exhibits a large bandgap of $\sim$1.42 eV that is only accessible in the optical regime, graphene behaves as a zero bandgap semiconductor, which means that both electron and hole carriers in this band structure are readily accessible with low energy excitations. Interesting physical phenomena and device applications based on these properties have been the subject of a large number of works in the recent literature.

Before moving on, we list some of the commonly used physical quantities that can be readily calculated from the E-k relation for reference in later chapters (Table 1.1). In the following, we will denote the density of electrons in 2D as n_0, while that in 3D will be denoted as n_{3D}, which we will occasionally use when a comparison to 3D metallic system is needed. All calculations will be presented at $T = 0$ K unless noted otherwise (see Sec. 2.3.2 for temperature dependent calculations). From n_0 or n_{3D}, the Fermi wavenumber k_{F} can be obtained via $n = \iint_{|\vec{k}|\leq k_{\mathrm{F}}} \frac{d^d\vec{k}}{(2\pi)^d} g$, where d is the dimensionality and g is the spin and valley degeneracy[2]. Fermi energy E_{F} is the value of $E(\vec{k})$ when $k = k_{\mathrm{F}}$. Fermi velocity v_{F} is calcu-

[2]Unless magnetic effects are considered, spin degeneracy is 2. Valley degeneracy depends on the band structure and symmetries of the specific material. For instance, GaAs/AlGaAs 2DEG has a valley degeneracy of 1, while Si inversion layer on (100) surface would have [3] a valley degeneracy of 2. 3D metals have a valley degeneracy of 1. g is the product of spin and

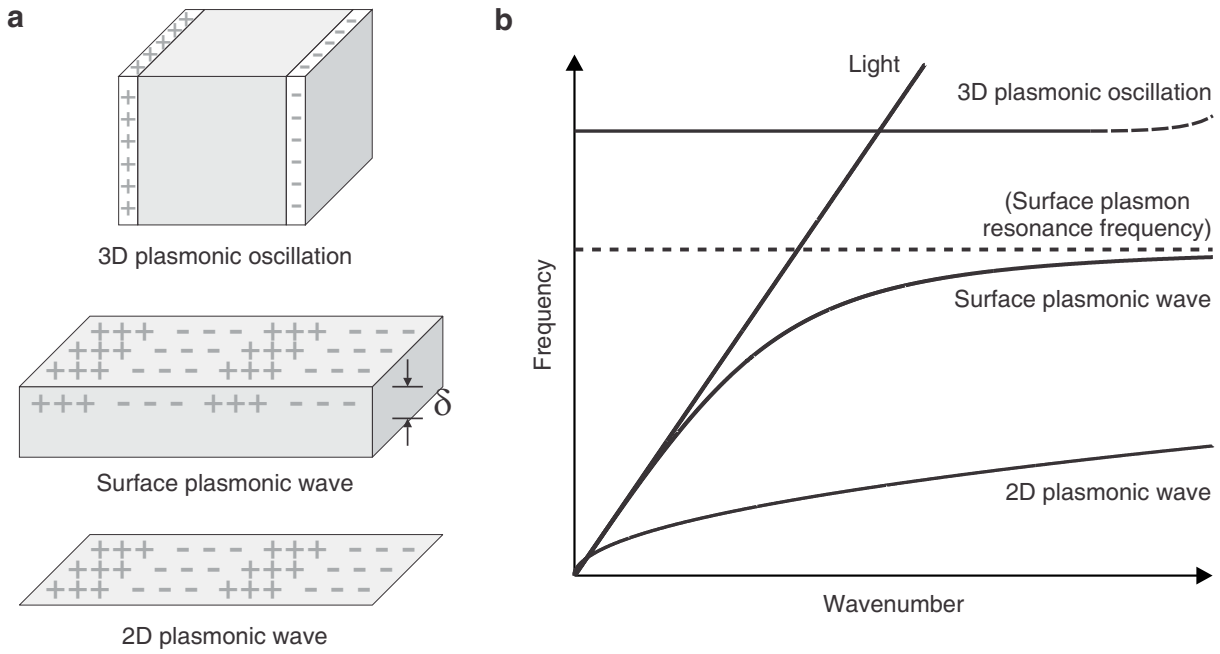

Figure 1.4: Comparison of 3D and 2D plasmonic waves. **a**, Illustration of the charge density distributions associated with 3D plasmonic oscillation, surface plasmonic wave on 3D metal, and 2D plasmonic wave on 2D conductor. **b**, Hypothetical dispersion curves for light wave, 3D plasmonic oscillation, surface plasmonic wave, and 2D plasmonic wave, capturing the essence of the difference between these entities.

lated as $v_{\mathrm{F}} = \frac{1}{\hbar}\frac{\partial E}{\partial k}\big|_{k=k_{\mathrm{F}}}$. Density of states $D(E)$, representing the number of states per unit area/volume with energy between E and $E + dE$, is obtained by $D(E) = \frac{\partial \tilde{n}}{\partial E}$, where $\tilde{n}(E) = \iint_{E(\vec{k})\leq E} \frac{d^d\vec{k}}{(2\pi)^d} g$.

1.2 Plasmonic Waves in 3D and 2D

Plasmonic oscillation in bulk 3D metal is a well known phenomenon characterized by the collective oscillation of electrons originating from the inertia of electrons and the Coulomb restoring force from the background positive charge (Fig. 1.4a, top). This can be easily explained in a classical picture as follows. When the electron gas in the metal is collectively shifted from the equilibrium position by a displacement $\vec{x}$, an electric field $\vec{E} = (en_{3D}/\epsilon_0)\vec{x}$ is created inside the metal. This leads to an equation of motion for a unit volume of electrons inside the metal, $(m^* n_{3D})\ddot{\vec{x}} = (-en_{3D})\vec{E}$, which can be rewritten as

$$\ddot{\vec{x}} = -\omega_{\mathrm{p}}^2 \vec{x}, \tag{1.3}$$

valley degeneracies.

where $\omega_{\mathrm{p}}^2 = n_{3D}e^2/m^*\epsilon_0$ is known as the plasma frequency. This describes a collective oscillation of the whole electron gas at angular frequency ω_{p}, which is typically in the optical frequency regime for 3D metals.

Assuming that the cross-sectional area of the metal is large enough, the electric field inside the metal is given by $\vec{E} = (en_{3D}/\epsilon_0)\vec{x}$ regardless of the size in the $\hat{x}$ direction. Hence, if we plot the dispersion relation for this plasmonic oscillation, it appears as a flat curve at frequency ω_{p} (Fig. 1.4b) unless higher order thermal [4] or quantum mechanical [5] effects are considered, which only happen at very large wavenumbers well above the regime of interest in this thesis. Note that we here only consider the purely longitudinal mode in the plasma that satisfies $\vec{k} \times \vec{E} = \omega\vec{B} = 0$ in Maxwell's equations. This mode is to be distinguished from the transverse electromagnetic mode in 3D metal above ω_{p} that has a dispersion relation of the form $\omega^2 = \omega_{\mathrm{p}}^2 + k^2c^2$, corresponding to a dielectric constant of $1 - (\omega_{\mathrm{p}}^2/\omega^2)$ for the metal.

At the surface of a 3D metal, a hybrid mode of electromagnetic wave and 3D plasmonic oscillation can propagate along the direction of the surface, because an oblique transverse-magnetic electromagnetic wave has a longitudinal electric field component along the direction of the surface, coupling the electromagnetic wave and 3D plasmonic oscillation in the vicinity of the surface (Fig. 1.4a, middle). Because the transverse electromagnetic wave cannot propagate *inside* the 3D metal below ω_{p} (it decays exponentially), this mode is contained at the skin of the metal in a frequency-dependent thickness δ known as the penetration depth. The dispersion of this wave (Fig. 1.4b) is similar to the light dispersion at low frequencies where most of the energy is propagated in electromagnetic form, while it deviates significantly away from the light dispersion when it approaches a specific frequency known as the surface plasmon resonance frequency, which is intimately related to ω_{p} (this dispersion relation will be discussed in greater detail in following chapters). The behavior near this frequency is particularly interesting as the wave propagates at a speed much slower than the speed of light, and thus can exhibit subwavelength confinement with proportionally reduced wavelength [6–9]. This has nucleated a great deal of research in photonics to create miniaturized optical structures.

Because the electrons in 2D conductors also possess inertia and the background positive charge in 2D still exerts Coulomb restoring force, 2D conductors can also support plasmonic waves (Fig. 1.4a, bottom). However, the plasmonic waves in 2D can no longer be explained in relation to the 3D plasmonic oscillation and requires a different formulation, which we set out to unfold in the coming chapters. The

dispersion of 2D plasmonic waves, shown in Fig. 1.4b, shows that these 2D plasmonic waves appear at much lower frequencies compared to surface plasmonic waves, typically at infrared and THz-GHz frequencies which is near or in the electronics regime. Furthermore, the 2D plasmons can achieve much greater subwavelength confinement [10–12] with their velocity being able to reach well below $c/100$ [11, 13].

By shaping the 2D conductor geometry with the standard fabrication technology and manipulating 2D plasmonic waves via reflection, interference, and coupling according to the geometry, a variety of ultra-subwavelength 2D plasmonic circuits and metamaterials, such as bandgap crystals, interferometers, resonant cavities, and negative refractive index structures, can be created [11, 13, 14] for GHz-THz and infrared integrated electronics with potential applications in imaging, large molecule spectroscopy, and sub-millimeter wave astronomy. The ultra-subwavelength confinement of these 2D plasmonic structures suggests exciting possibilities for sub-diffraction-limit imaging, near-field operation, and drastic miniaturization.

In the following chapters, we will first elucidate the unique behavioral characteristics of plasmonic waves in 2D conducting media and their physical origin (Chapter 2). Then we will delineate how 2D plasmonic waves can be engineered to build functional circuits and metamaterials, implemented using semiconductor heterojunction 2DEG (Chapter 3) and graphene (Chapter 4) as the 2D conducting media.

Chapter 2

Characteristics of 2D Plasmonic Waves

Perturbation of the equilibrium electron density distribution in a solid-state conductor, whether 3D or 2D, results in Coulomb restoring force, which drives local electrons back and forth collectively to propel a plasmonic wave[1]. The defining energetic component of a plasmonic wave is the kinetic energy of the collectively oscillating electrons, which largely accounts for the plasmons' behavioral difference from light waves, in particular, the reduced plasmonic velocity and subwavelength confinement. The kinetic energy is far more strongly pronounced in 2D plasmonic waves than in 3D bulk metal surface plasmonic waves [11, 14]. Consequently, the behavior of 2D plasmons diverges even more significantly from light waves than that of 3D surface plasmons does; for example, and notably, 2D plasmons can achieve a significantly lower velocity thus a much greater subwavelength confinement than 3D surface plasmons. This chapter explicates the origin of the unique behaviors of 2D plasmonic waves in comparison to 3D bulk metal surface plasmonic waves.

2.1 Transmission Line Model for 2D Plasmonic Medium

As will be discussed shortly, the kinetic energy of the collectively oscillating electrons in a 2D plasmonic wave can be modeled using kinetic inductance of non-magnetic origin [11, 14]. On the other hand, the electric potential energy associated with the Coulomb restoring force that drives local electrons into the plasmonic oscillation can be modeled using electrical capacitance. Besides the Coulomb restoring force,

[1] Large portions of this chapter are derived from a number of papers in publication written by the author [14–17].

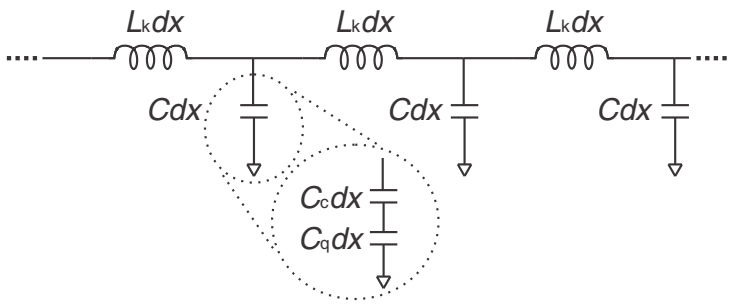

Figure 2.1: Lossless transmission line model of a 2D plasmonic medium [16]. If the medium is gated, the gate serves as the ground. For an ungated medium, the ground is the potential of the free space far away from the 2D medium. dx: infinitesimal segment length of the 2D plasmonic medium.

electron degeneracy pressure serves as another restoring mechanism upon the disturbance of the equilibrium electron density distribution, and this effect can be modeled using quantum capacitance [10,18,19]. Then the 2D plasmonic medium can be modeled as a transmission line consisting of distributed kinetic inductance $L_{\rm k}$ per unit length and distributed capacitance C per unit length (Fig. 2.1) [11,14,20], where the effects of the two capacitances have been combined into one effective capacitance $C = \left(C_{\rm c}^{-1} + C_{\rm q}^{-1}\right)^{-1}$ with $C_{\rm c}$ and $C_{\rm q}$ being classical (geometric, electric) and quantum capacitances, respectively. The effect of magnetic self-inductance, which is usually orders of magnitude smaller than that of kinetic inductance in 2D conductors as will be discussed in greater detail in Chapter 3 and 4, is omitted in this model. We first consider the lossless case for simplicity. This plasmonic transmission line differs from the standard electromagnetic transmission line in that the latter employs magnetic inductance instead of kinetic inductance. The plasmonic velocity is then $v_{\rm p} = 1/\sqrt{L_{\rm k}C}$, which corresponds to the plasmonic dispersion relation.

2.1.1 Calculation of $L_{\rm k}$

The kinetic inductance $L_{\rm k}$ modeling the kinetic energy of the collectively oscillating electrons can be calculated for each 2D plasmonic medium. In this section, we first calculate $L_{\rm k}$ for a 2D plasmonic medium consisting of massive electrons as would be appropriate for a semiconductor heterojunction 2DEG, then calculate $L_{\rm k}$ for a medium consisting of massless electrons as would be appropriate for graphene, and then provide a generic calculation insensitive to the individual electron dispersion.

2.1.1.1 Massive Individual Electrons

$L_{\rm k}$ in a 2D conductor where individual electrons have finite effective mass m^*, such as in GaAs/AlGaAs 2DEG, can be calculated in a classical picture [14]. Let a time-dependent electric potential $V(t)$ be applied along a strip of the 2D conductor (width W and length l) to induce an electric field $V(t)/l$. Here the length l is chosen so short that the electric field does not exhibit a spatial variation; this is not a limiting assumption, as the goal is to derive the kinetic inductance per unit length. Inertial accelerations occur, for which Newton's equation of motion for an electron is

$$-e\frac{V}{l} = m^*\frac{dv}{dt} \tag{2.1}$$

(v: electron velocity). This translates to $-e\frac{V}{l} = i\omega m^* v$ in the frequency domain. From this and by noting that the current due to the electrons' motion is $I = -n_0 evW$ (n_0: conduction electron density per unit area), the 2D conductor's impedance is obtained:

$$\frac{V}{I} = i\omega\frac{m^*}{n_0 e^2}\frac{l}{W}. \tag{2.2}$$

This is an inductive impedance of kinetic origin, with the inductance value per unit length given by

$$L_{\rm k} = \frac{m^*}{n_0 e^2}\frac{1}{W}. \tag{2.3}$$

The total kinetic energy of the accelerating electrons is intimately linked to the kinetic inductance. With the velocity v of an electron at a given time, the total kinetic energy $E_{\rm total}$ of the electrons in the 2D conductor strip is expressed as $E_{\rm total} = \frac{1}{2}m^* v^2 n_0 W l$. Since the total current is $I = -n_0 evW$, we can write

$$E_{\rm total} = \frac{1}{2}\left(L_{\rm k} l\right) I^2, \tag{2.4}$$

where $L_{\rm k} l$ is the total kinetic inductance of the 2D conductor strip. Eq. (2.4) is analogous to the energy of a magnetic inductor with current I given by $\frac{1}{2}(\text{inductance})I^2$.

The intuitive derivation above, however, does not provide the correct picture of the electrons following Fermi-Dirac statistics, even though the correct expression Eq. (2.3) was obtained. Alternatively, we

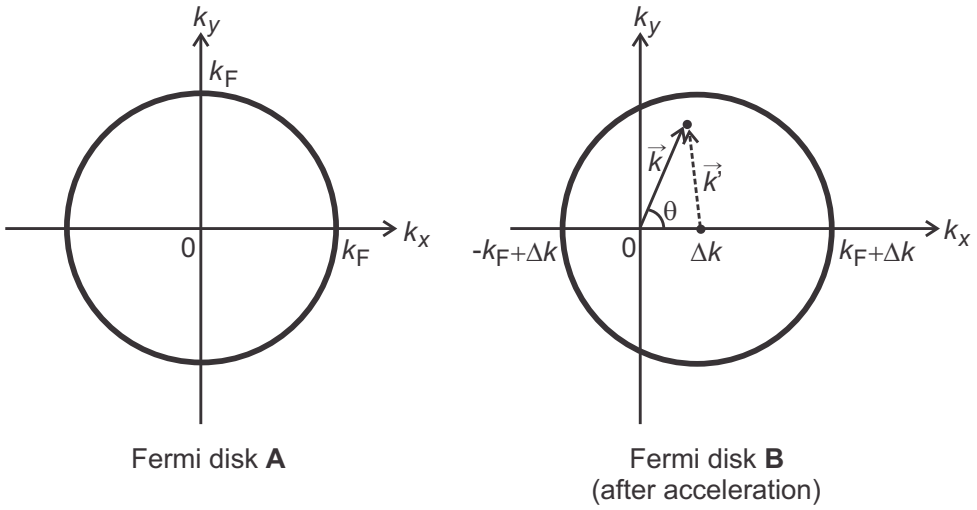

Figure 2.2: Shift of the Fermi disk in k-space in response to an electric field [14].

can instead calculate $L_{\rm k}$ by first evaluating the total kinetic energy $E_{\rm total}$ and current I in the k-space (k: electron wavenumber) and then relating them through the energy-current relation [14]. With the electric field applied along the length of the 2D conductor strip to which direction we assign a negative x-axis, the 2D Fermi disk with radius $k_{\rm F}$ whose center originally lies at the k-space origin (Fig. 2.2, **A**) shifts towards the positive k_x-axis, increasing the total kinetic energy and producing a current I. Fig. 2.2, **B** shows the Fermi disk shift by $\Delta k \ll k_{\rm F}$. The total kinetic energy increase per unit area is

$$\frac{E_{\rm total}}{Wl} = \iint\limits_{\mathbf{B}} \frac{d^2\vec{k}}{(2\pi)^2} gE(\vec{k}) - \iint\limits_{\mathbf{A}} \frac{d^2\vec{k}}{(2\pi)^2} gE(\vec{k}), \tag{2.5}$$

where g accounts for spin and valley degeneracy (*e.g.*, $g = 2$ for GaAs/AlGaAs 2DEG) and $E(\vec{k}) = \frac{\hbar^2 k^2}{2m^*}$ is the energy of a single electron whose wavevector is $\vec{k}$. The first integration is computed as below, where $\vec{k'} = \vec{k} - \Delta k \hat{e}_{k_x}$ ($\hat{e}_{k_x}$: unit vector along the k_x direction) is with reference to Fig. 2.2:

$$\iint\limits_{\mathbf{B}} \frac{d^2\vec{k}}{(2\pi)^2} gE(\vec{k}) = \frac{g\hbar^2}{8\pi^2 m^*} \iint\limits_{\mathbf{B}} d^2\vec{k'} \left|\vec{k'} + \Delta k \hat{e}_{k_x}\right|^2 = \frac{g\hbar^2}{8\pi^2 m^*} \iint\limits_{\mathbf{B}} d^2\vec{k'} \left[k'^2 + 2k'_x \Delta k + (\Delta k)^2\right]. \tag{2.6}$$

This contains three terms. By inspection of Fig. 2.2, the first term is simply the second integration of Eq. (2.5), the total kinetic energy per unit area corresponding to the Fermi disk **A**. The second term vanishes due to the odd symmetry of k'_x within the disk **B**. Therefore Eq. (2.5) reduces to (the integration

is performed at $T = 0$ K for simplicity)

$$\frac{E_{\rm total}}{Wl} = \frac{g\hbar^2}{8\pi^2 m^*}(\Delta k)^2 \iint\limits_{\mathbf{B}} d^2\vec{k'} = \frac{g\hbar^2 k_{\rm F}^2}{8\pi m^*}(\Delta k)^2. \tag{2.7}$$

The total current I is calculated as

$$I = W \iint\limits_{\mathbf{B}} \frac{d^2\vec{k}}{(2\pi)^2} g e v_x(\vec{k}), \tag{2.8}$$

where $v_x(\vec{k}) = \hbar k_x/m^*$ is the x-component of the velocity of an electron with wavevector $\vec{k}$. Thus,

$$I = \frac{Wge\hbar}{4\pi^2 m^*} \iint\limits_{\mathbf{B}} d^2\vec{k} k_x = \frac{Wge\hbar}{4\pi^2 m^*} \int_0^{2\pi} d\theta \cos\theta \int_0^{k_{\rm F}+\Delta k\cos\theta} k^2 dk, \tag{2.9}$$

where the distance between the origin and the edge of the disk $\mathbf{B}$ at angle θ is approximated as $k_{\rm F} + \Delta k \cos\theta$, given $\Delta k \ll k_{\rm F}$. By performing the integration up to the first order of Δk (the lowest surviving order), we obtain

$$I = \frac{Wge\hbar k_{\rm F}^2}{4\pi m^*}\Delta k. \tag{2.10}$$

By eliminating Δk using Eqs. (2.7) and (2.10), and noting $k_{\rm F}^2 = 4\pi n_0/g$, we get

$$E_{\rm total} = \frac{1}{2}\left[\frac{m^*}{n_0 e^2}\frac{l}{W}\right] I^2, \tag{2.11}$$

which is identical to Eq. (2.4) with $L_{\rm k}$ given by Eq. (2.3).

We note that this energy-current relation links the voltage V, which produces the electric field and moves the electrons, and the current I in exactly the same *mathematical* manner as magnetic inductance links voltage across it and current through it, *viz.*, $V = (L_{\rm k} l)\frac{dI}{dt}$. To demonstrate, one can note that $E_{\rm total} = \int_{t_0}^{t} VIdt = \frac{1}{2}(L_{\rm k} l)I^2$ (t_0 is when the electron gas is at rest) and differentiate this equation with respect to time t. In fact, just as magnetic inductance represents the reluctance of magnetic flux to change, the kinetic inductance represents the 'inertial' reluctance of the collective momentum to change.

2.1.1.2 Massless Individual Electrons

For a medium consisting of massless electrons, the classical picture of individual electrons moving according to the Newton's equation of motion cannot be applied. Nonetheless, the derivation of Eqs. (2.5) - (2.11) can still be used to show that the medium consisting of massless electrons also exhibits an inductive impedance. This implies that the collection of massless electrons behaves as if it collectively possesses an inertial mass, which will be seen in greater detail in Chapter 4. In this section, we focus on deriving the kinetic inductance of such a medium.

The situation assumed to calculate the kinetic inductance is identical to what was seen in Fig. 2.2. The total kinetic energy increase per unit area is calculated as Eq. (2.5), but with $E(\vec{k}) = \hbar v_{\mathrm{F}} k$ as would be appropriate for a 2D medium consisting of massless electrons, such as graphene ($g = 4$ for graphene due to the two degenerate valleys). The integration of Eq. (2.6) then becomes

$$\iint\limits_{\mathbf{B}} \frac{d^2\vec{k}}{(2\pi)^2} g E(\vec{k}) = \frac{g\hbar v_{\mathrm{F}}}{4\pi^2} \iint\limits_{\mathbf{B}} d^2\vec{k'} \left|\vec{k'} + \Delta k \hat{e}_{k_x}\right| = \frac{g\hbar v_{\mathrm{F}}}{4\pi^2} \iint\limits_{\mathbf{B}} d^2\vec{k'} \left[k'^2 + 2k'_x \Delta k + (\Delta k)^2\right]^{1/2}. \tag{2.12}$$

By expanding the integrand to the second order of Δk, which is the lowest surviving order as will be seen shortly, we obtain

$$\iint\limits_{\mathbf{B}} \frac{d^2\vec{k}}{(2\pi)^2} g E(\vec{k}) = \frac{g\hbar v_{\mathrm{F}}}{4\pi^2} \iint\limits_{\mathbf{B}} d^2\vec{k'} \left[k' + \frac{k'_x}{k'}\Delta k + \frac{k'^2_y}{2k'^3}(\Delta k)^2\right]. \tag{2.13}$$

Since the first term on the right hand side is identical to the second integration of Eq. (2.5) and the second term vanishes due to the odd symmetry of k'_x within the disk **B**, Eq. (2.5) is simplified to

$$\frac{E_{\mathrm{total}}}{Wl} = \frac{g\hbar v_{\mathrm{F}}}{4\pi^2} \iint\limits_{\mathbf{B}} d^2\vec{k'} \frac{k'^2_y}{2k'^3}(\Delta k)^2 = \frac{g\hbar v_{\mathrm{F}} k_{\mathrm{F}}}{8\pi}(\Delta k)^2. \tag{2.14}$$

This quadratic dependency of E_{total} on small enough Δk arises because E_{total} as a function of Δk must assume a smooth extremum (minimum) when $\Delta k = 0$, *i.e.*, at the collective ground state.

The total current I is calculated as Eq. (2.8), but with $v_x(\vec{k}) = v_{\mathrm{F}} \cos\theta$. Thus,

$$I = \frac{W g e v_{\mathrm{F}}}{4\pi^2} \iint_{\mathbf{B}} d^2\vec{k} \cos\theta = \frac{W g e v_{\mathrm{F}}}{4\pi^2} \int_0^{2\pi} d\theta \cos\theta \int_0^{k_{\mathrm{F}} + \Delta k \cos\theta} k dk, \tag{2.15}$$

where the distance between the origin and the edge of the disk **B** at angle θ is approximated as $k_{\mathrm{F}} + \Delta k \cos\theta$, given $\Delta k \ll k_{\mathrm{F}}$. By performing the integration up to the first order of Δk (the lowest surviving order), we obtain

$$I = \frac{W g e v_{\mathrm{F}} k_{\mathrm{F}}}{4\pi} \Delta k. \tag{2.16}$$

By eliminating Δk using Eqs. (2.14) and (2.16), we obtain

$$E_{\mathrm{total}} = \frac{1}{2} \left[\frac{4\pi\hbar}{g e^2 v_{\mathrm{F}} k_{\mathrm{F}}} \frac{l}{W} \right] I^2, \tag{2.17}$$

from which we identify the kinetic inductance per unit length as [16, 21]

$$L_{\mathrm{k}} = \frac{4\pi\hbar}{g e^2 v_{\mathrm{F}} k_{\mathrm{F}}} \frac{1}{W}. \tag{2.18}$$

We note that if Eq. (2.3) is expressed in terms of k_{F} and v_{F} by substituting $m^* = \hbar k_{\mathrm{F}} / v_{\mathrm{F}}$, Eq. (2.3) is identical to Eq. (2.18).

2.1.1.3 Arbitrary Isotropic Electron Dispersion

The previous derivations assumed a quadratic (massive) or linear (massless) individual electron dispersion. However, it can be shown that E_{total} is proportional to $(\Delta k)^2$ for any arbitrary isotropic dispersion, and hence the electron gas collectively exhibits kinetic inductance regardless of the individual electron's character in the medium.

This can be seen by focusing on the behavior of the electron gas near the Fermi surface. As seen in Fig. 2.3, the shift of the Fermi disk results in electrons added in the shaded region **C** and removed in region **D**. Because $\Delta k \ll k_{\mathrm{F}}$, the arbitrary isotropic dispersion within the shaded regions can be approximated as

$$E(\vec{k}) = E(k_{\mathrm{F}}) + \hbar v_{\mathrm{F}} (k - k_{\mathrm{F}}), \tag{2.19}$$

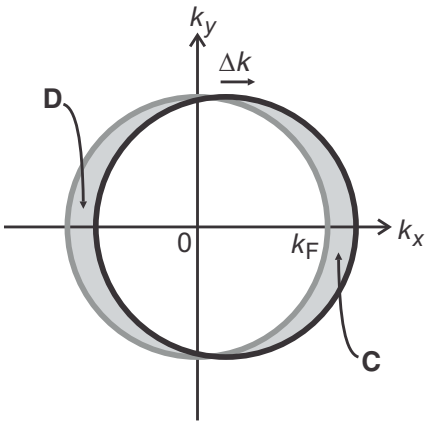

Figure 2.3: Shift of the Fermi disk, focusing on the behavior near the Fermi surface.

where $(k-k_{\rm F})$ remains at or smaller than the first order of Δk. With reference to Fig. 2.3, we can rewrite the total kinetic energy increase per unit area in Eq. (2.5) as

$$\frac{E_{\rm total}}{Wl} = \iint\limits_{\mathbf{C}} \frac{d^2\vec{k}}{(2\pi)^2} gE(\vec{k}) - \iint\limits_{\mathbf{D}} \frac{d^2\vec{k}}{(2\pi)^2} gE(\vec{k}), \tag{2.20}$$

where

$$\begin{aligned} \iint\limits_{\mathbf{C}} d^2\vec{k} &= \int_0^{\pi} d\theta \int_{k_{\rm F}}^{k_{\rm F}+\Delta k \sin\theta} dk\big(k - \frac{\Delta k}{2}\sin\theta\big), \\ \iint\limits_{\mathbf{D}} d^2\vec{k} &= \int_0^{\pi} d\theta \int_{k_{\rm F}-\Delta k \sin\theta}^{k_{\rm F}} dk\big(k + \frac{\Delta k}{2}\sin\theta\big), \end{aligned} \tag{2.21}$$

given $\Delta k \ll k_{\rm F}$ (for convenience in notations, θ has been rotated by $-\frac{\pi}{2}$ for **C** and $\frac{\pi}{2}$ for **D**). The $\mp\frac{\Delta k}{2}\sin\theta$ part is necessary to ensure that the total number of electrons is conserved during the Fermi disk shift, *i.e.*,

$$\iint\limits_{\mathbf{C}} d^2\vec{k} = \iint\limits_{\mathbf{D}} d^2\vec{k} = 2k_{\rm F}\Delta k. \tag{2.22}$$

The k-integrations in Eq. (2.20) can now be readily performed:

$$\begin{aligned}\int_{k_{\rm F}}^{k_{\rm F}+\Delta k\sin\theta} dk\big(k-\frac{\Delta k}{2}\sin\theta\big)E(k) &= \frac{1}{3}\hbar v_{\rm F}\big(3k_{\rm F}^2\Delta k\sin\theta+3k_{\rm F}(\Delta k)^2\sin^2\theta+(\Delta k)^3\sin^3\theta\big)\\ &+\frac{1}{2}\big(E(k_{\rm F})-\hbar v_{\rm F}k_{\rm F}-\frac{1}{2}\hbar v_{\rm F}\Delta k\sin\theta\big)\big(2k_{\rm F}\Delta k\sin\theta+(\Delta k)^2\sin^2\theta\big)\\ &-\frac{1}{2}(E(k_{\rm F})-\hbar v_{\rm F}k_{\rm F})(\Delta k)^2\sin^2\theta,\end{aligned} \tag{2.23}$$

$$\begin{aligned}\int_{k_{\rm F}-\Delta k\sin\theta}^{k_{\rm F}} dk\big(k+\frac{\Delta k}{2}\sin\theta\big)E(k) &= \frac{1}{3}\hbar v_{\rm F}\big(3k_{\rm F}^2\Delta k\sin\theta-3k_{\rm F}(\Delta k)^2\sin^2\theta+(\Delta k)^3\sin^3\theta\big)\\ &+\frac{1}{2}\big(E(k_{\rm F})-\hbar v_{\rm F}k_{\rm F}+\frac{1}{2}\hbar v_{\rm F}\Delta k\sin\theta\big)\big(2k_{\rm F}\Delta k\sin\theta-(\Delta k)^2\sin^2\theta\big)\\ &+\frac{1}{2}(E(k_{\rm F})-\hbar v_{\rm F}k_{\rm F})(\Delta k)^2\sin^2\theta.\end{aligned} \tag{2.24}$$

Combining the two integrations, Eq. (2.20) is greatly simplified to yield

$$\frac{E_{\rm total}}{Wl}=\frac{g}{4\pi^2}\int_0^\pi d\theta\hbar v_{\rm F}k_{\rm F}(\Delta k)^2\sin^2\theta=\frac{g\hbar v_{\rm F}k_{\rm F}}{8\pi}(\Delta k)^2. \tag{2.25}$$

We note that this result is identical to Eq. (2.14).

The total current I is calculated in a manner similar to Eq. (2.15), where we can use the approximation $v_x(\vec{k})=v_{\rm F}\cos\theta$ as $\Delta k\ll k_{\rm F}$ and we limit ourselves to regions **C** and **D** near the Fermi surface. Thus, (again, θ has been rotated by $-\frac{\pi}{2}$ for **C** and $\frac{\pi}{2}$ for **D**)

$$\begin{aligned}I=&\frac{Wge}{4\pi^2}\Big(\iint\limits_{\mathbf{C}} d^2\vec{k}v_x(\vec{k})-\iint\limits_{\mathbf{D}} d^2\vec{k}v_x(\vec{k})\Big)\\ =&\frac{Wgev_{\rm F}}{4\pi^2}\int_0^\pi d\theta\Big[\int_{k_{\rm F}}^{k_{\rm F}\Delta k\sin\theta} dk\big(k-\frac{\Delta k}{2}\sin\theta\big)\sin\theta+\int_{k_{\rm F}-\Delta k\sin\theta}^{k_{\rm F}} dk\big(k+\frac{\Delta k}{2}\sin\theta\big)\sin\theta\Big],\end{aligned} \tag{2.26}$$

which can be readily computed to obtain

$$I=\frac{Wgev_{\rm F}}{4\pi^2}\int_0^\pi d\theta 2k_{\rm F}\Delta k\sin^2\theta=\frac{Wgev_{\rm F}k_{\rm F}}{4\pi}\Delta k. \tag{2.27}$$

We note that this result is identical to Eq. (2.16). Since both Eq. (2.25) and Eq. (2.27) are identical to

Eq. (2.14) and Eq. (2.16), we arrive at the same expression for the kinetic inductance per unit length,

$$L_{\mathrm{k}} = \frac{4\pi\hbar}{ge^2 v_{\mathrm{F}} k_{\mathrm{F}}} \frac{1}{W}. \tag{2.28}$$

As stated earlier, the quadratic dependency of E_{total} on small enough Δk, and the kinetic inductance L_{k} that follows, arise because E_{total} as a function of Δk must assume a smooth extremum (minimum) when $\Delta k = 0$, *i.e.*, at the collective ground state. The expression for L_{k} is insensitive to the shape of the single electron dispersion as the collective behavior effectively happens only in the vicinity of the Fermi surface, within which only the slope of the dispersion, *i.e.*, $\hbar v_{\mathrm{F}}$, is relevant (Eq. (2.19)).

2.1.2 Calculation of C

The per-unit-length distributed capacitance C in the 2D plasmonic transmission line (Fig. 2.1) contains two components, namely, the geometric capacitance C_{c} and the quantum capacitance C_{q}. In terms of energy, these two components can be seen as the amount of energy added to the system by adding a unit charge to the 2D conductor, with the energy originating from the electric fields surrounding the 2D conductor or from the kinetic energy of the electrons being added, respectively. Here we discuss how these capacitances can be obtained for a 2D conductor.

2.1.2.1 Geometric Capacitance C_{c}

The per-unit-length geometric capacitance C_{c}, which models the Coulomb restoring force in the plasmonic wave, depends on the surroundings of the plasmonic medium. For example, if a 2D conductor strip with width W has no other conductors nearby, C_{c} for a given plasmonic wavenumber k_{p} is given by [22, 23]

$$C_{\mathrm{c}} = 2\kappa\epsilon_0 k_{\mathrm{p}} W, \tag{2.29}$$

where κ is the effective dielectric constant of the surrounding medium. This is obtained by integrating the energy density stored in electric fields around the 2D conductor assuming a sinusoidal charge density distribution at a plasmonic wavenumber k_{p}.

If an external conductor is proximate to the 2D conductor, C_{c} is altered. A case of particular interest is a gated 2D conductor. If the separation d between the gate and the 2D conductor is much smaller than

the plasmonic wavelength (*i.e.*, $k_{\mathrm{p}}d \ll 1$), C_{c} becomes the parallel plate capacitance per unit length,

$$C_{\mathrm{c}} = \kappa\epsilon_0 \frac{W}{d}. \tag{2.30}$$

2.1.2.2 Quantum Capacitance C_{q}

The per-unit-length quantum capacitance C_{q} represents the increase in total kinetic energy of the electron gas by adding a unit charge to the system (in contrast, L_{k} represented the total kinetic energy increase due to the movement of charges while the total number of electrons was fixed). Therefore, this capacitance is directly proportional to the density of states $D(E)$ at the Fermi surface and is given by $C_{\mathrm{q}} = e^2 D(E_{\mathrm{F}})W$ [24]. In principle, this capacitance arises from Pauli's exclusion principle as any electron added into the system needs to occupy energy levels above the Fermi level.

The quantum capacitance for a 2D electron gas consisting of massive electrons is given by [3, 18] (calculated at $T = 0$ K)

$$C_{\mathrm{q}} = \frac{gm^* e^2}{2\pi\hbar^2} W, \tag{2.31}$$

where g accounts for spin and valley degeneracy as in Eq. (2.5). The quantum capacitance for a 2D electron gas consisting of massless electrons is given by [3, 19, 24] (calculated at $T = 0$ K)

$$C_{\mathrm{q}} = \frac{ge^2 E_{\mathrm{F}}}{2\pi(\hbar v_{\mathrm{F}})^2} W, \tag{2.32}$$

where $E_{\mathrm{F}} = \hbar v_{\mathrm{F}} k_{\mathrm{F}}$. We note that Eqs. (2.31) and (2.32) can both be written by the same expression as

$$C_{\mathrm{q}} = \frac{ge^2 k_{\mathrm{F}}}{2\pi\hbar v_{\mathrm{F}}} W. \tag{2.33}$$

2.2 Dispersion Relation of 2D Plasmonic Waves

In the previous section, we have obtained the constituent components of the 2D plasmonic transmission line model (Fig. 2.1), namely, the kinetic inductance L_{k}, the geometric capacitance C_{c}, and the quantum capacitance C_{q}. Using these quantities, transmission line theory [25] can be applied to show the interesting properties of 2D plasmonic waves.

The potential energy $V(x)$ and current $I(x)$ in the transmission line model of Fig. 2.1 obey wave equations

$$\begin{aligned} \frac{d^2V(x)}{dx^2} - \gamma^2 V(x) = 0, \\ \frac{d^2I(x)}{dx^2} - \gamma^2 I(x) = 0, \end{aligned} \tag{2.34}$$

where

$$\gamma \equiv \alpha + ik_{\mathrm{p}} = i\omega\sqrt{L_{\mathrm{k}}C} \tag{2.35}$$

is known as the complex propagation constant, with α representing the attenuation (α=0 for the lossless transmission line model of Fig. 2.1) and k_{p} representing the plasmonic wavenumber (α and k_{p} are real by definition). The wave equations allow general solutions of the form

$$\begin{aligned} V(x) = V_0^+ e^{-\gamma x} + V_0^- e^{\gamma x}, \\ I(x) = I_0^+ e^{-\gamma x} + I_0^- e^{\gamma x}, \end{aligned} \tag{2.36}$$

where the $e^{-\gamma x}$ term represents wave propagation in the $+x$ direction and the $e^{\gamma x}$ term represents wave propagation in the $-x$ direction. The wave amplitudes $V_0^{\pm}$ and $I_0^{\pm}$ satisfy

$$Z_0 \equiv \sqrt{\frac{L}{C}} = \frac{V_0^+}{I_0^+} = -\frac{V_0^-}{I_0^-}, \tag{2.37}$$

where Z_0 is known as the characteristic impedance of the transmission line.

In particular, Eq. (2.35) reveals the dispersion relation of 2D plasmonic waves, the relationship between the wavenumber k_{p} and frequency ω. It is worthwhile to consider 2D plasmonic waves in two notable and practically useful cases, namely, stand-alone ('ungated') 2D conductor and proximately gated 2D conductor, which we will compare to surface plasmonic waves on 3D metals (Fig. 2.4). We consider both massive and massless electrons, but since their kinetic inductance L_{k} and quantum capacitance C_{q} can both be written by the same expressions Eqs. (2.28) and (2.33), their behaviors can be described together.

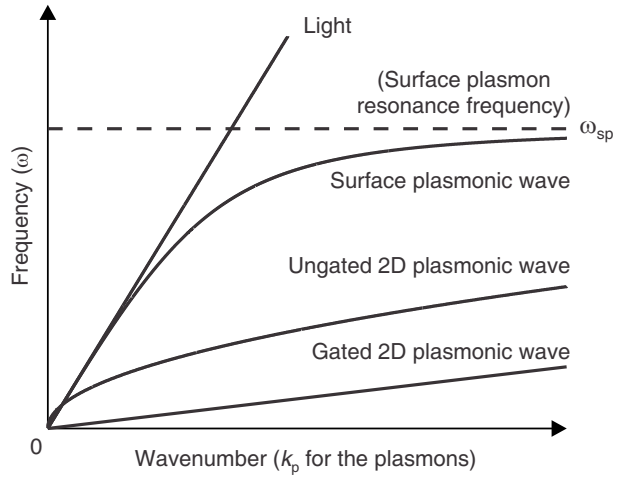

Figure 2.4: Qualitative comparison of dispersion relations for light, surface plasmonic wave on 3D metal, ungated 2D plasmonic wave, and gated 2D plasmonic wave [15]. The frequency scales of the dispersions for 2D plasmonic waves were exaggerated to allow visual comparison of their shapes with surface plasmonic waves. With more realistic values of E_{F} assumed for the metal and the 2D conductor on which these waves propagate, the 2D plasmonic waves typically appear at much lower frequencies (tens of GHz to tens of THz range) as opposed to surface plasmonic waves that appear at optical frequencies (hundreds of THz).

2.2.1 Ungated 2D Plasmonic Waves

2.2.1.1 Dispersion Relation

We first consider the stand-alone 2D conductor with no other conductors nearby. By combining Eqs. (2.28), (2.29), (2.33), and (2.35), we arrive at the following dispersion relation:

$$\omega = v_{\mathrm{F}} k_{\mathrm{F}} \sqrt{\frac{ge^2}{8\pi\kappa\epsilon_0\hbar v_{\mathrm{F}}}\left(\frac{k_{\mathrm{p}}}{k_{\mathrm{F}}}\right) + \frac{1}{2}\left(\frac{k_{\mathrm{p}}}{k_{\mathrm{F}}}\right)^2}. \tag{2.38}$$

Inside the square-root, the first term originates from the geometric capacitance and the second term originates from the quantum capacitance. In order to determine which term has a more dominating effect, we can rewrite the coefficient of the first term as

$$\left(\frac{g}{2\kappa}\right)\left(\frac{e^2}{4\pi\epsilon_0\hbar c}\right)\left(\frac{c}{v_{\mathrm{F}}}\right). \tag{2.39}$$

The first factor is on the order of 1 as most practical substrates lead to an effective dielectric constant κ smaller than 10. The second factor is the fine structure constant, a fundamental dimensionless constant of approximate numerical value $1/137$. The third factor is the ratio of the speed of light to the Fermi

velocity of the electron gas. For typical semiconductor heterojunction based 2DEGs, the Fermi velocity is $2 \sim 3$ orders of magnitude smaller than the speed of light [11, 13, 14, 20]. For graphene, the Fermi velocity is known to be ~ 300 times smaller than the speed of light [3, 26]. Therefore, for most practical situations, the coefficient of the first term in the square-root of Eq. (2.38) is on the order of 1. This means that for k_{p} smaller than k_{F}, the first term dominates, while for k_{p} larger than k_{F}, the second term dominates.

Often, one operates the 2D plasmonic waves at low enough frequencies where k_{p} is much smaller than k_{F}. In this long-wavelength regime, the effect of the quantum capacitance is negligible and the dispersion relation of Eq. (2.38) is approximated as [3, 11, 13, 14, 20, 26–28]

$$\omega = \sqrt{\frac{ge^2 v_{\mathrm{F}} k_{\mathrm{F}}}{8\pi\kappa\epsilon_0\hbar} k_{\mathrm{p}}}. \tag{2.40}$$

This characteristic square-root shape of the dispersion is displayed in Fig. 2.4.

2.2.1.2 Comparison to RPA Calculations

The dispersion relation for 2D plasmonic waves can also be obtained from a more general approach of random phase approximation (RPA). The foregoing formalism only considered collective electron movements within a single band (intra-band). However, at high enough excitation frequencies or wavenumbers, creation of electron-hole pairs within the band (Landau damping) or transitions between different bands (inter-band effects) can happen [27] which cannot be described by the simple model of Fig. 2.1. Description of 2D plasmonic waves using RPA is more general in that all of those effects can be handled within its formalism.

The aforementioned effects however only take place in the large-k_{p} regime ($k_{\mathrm{p}} > k_{\mathrm{F}}$). In frequency or wavenumber regimes where such effects do not happen, the dispersion relation derived using RPA reduces to Eq. (2.38), which we show below. While being limited in frequency range, the approach presented in the foregoing sections is beneficial as: it provides more physical insight as to how the energy is distributed into kinetic (inductive and capacitive) and electric (capacitive) energies in the 2D plasmonic wave; it is more amenable for the purpose of engineering devices utilizing 2D plasmonic waves; the frequency regime where the aforementioned damping mechanisms do not happen is the practically useful

regime.

In the random phase approximation, the longitudinal dielectric function in the 2D conductor can be written as

$$\epsilon_{||}(\vec{q},\omega) = 1 - \frac{e^2}{2\kappa\epsilon_0 q}\Pi_0(\vec{q},\omega), \tag{2.41}$$

where $e^2/2\kappa\epsilon_0 q$ is the 2D Fourier transform of the Coulomb interaction and $\Pi_0(\vec{q},\omega)$ is the 2D polarizability, which can include both contributions from intra-band and inter-band effects. As our interest here is to show the equivalence of the RPA approach and the transmission line model approach at small wavevectors, we only use the intra-band polarizability. For small wavevectors (*i.e.*, $v_{\rm F}q < \omega$), $\Pi_{0,\rm intra}(\vec{q},\omega)$ can be expressed in the following form [22]:

$$\Pi_{0,\rm intra}(\vec{q},\omega) = \frac{g v_{\rm F} k_{\rm F} q^2}{4\pi\hbar}\frac{1}{\omega^2 - v_{\rm F}^2 q^2/2}, \tag{2.42}$$

where $T \to 0$ K and $\tau \to \infty$ were assumed. Plasmon dispersion is found by obtaining wavevectors $\vec{q}$ that satisfy $\epsilon_{||}(\vec{q},\omega) = 0$, where we identify the real part of those values of q as $k_{\rm p}$. Thus, we obtain

$$\omega = \sqrt{\frac{g e^2 v_{\rm F} k_{\rm F}}{8\pi\kappa\epsilon_0\hbar}q + \frac{1}{2}v_{\rm F}^2 q^2}, \tag{2.43}$$

which is identical to Eq. (2.38) if we replace $q \to k_{\rm p}$. Therefore, we conclude that the transmission line model approach is consistent with the more general RPA description of 2D plasmons.

2.2.2 Gated 2D Plasmonic Waves

The dispersion relation for the proximately gated (*i.e.*, $k_{\rm p}d \ll 1$) 2D plasmonic wave can be found by combining Eqs. (2.28), (2.30), (2.33), and (2.35):

$$\omega = \sqrt{\frac{g e^2 k_{\rm F} d}{4\pi\kappa\epsilon_0\hbar v_{\rm F}} + \frac{1}{2}v_{\rm F}k_{\rm p}}. \tag{2.44}$$

This characteristic linear shape of the dispersion is displayed in Fig. 2.4. As in the ungated case, the first term in the square-root originates from the geometric capacitance and the second term originates from the quantum capacitance. In order to determine which term has a more dominating effect, we can rewrite

the first term in the square-root as

$$\left(\frac{g}{\kappa}\right)\left(\frac{e^2}{4\pi\epsilon_0\hbar c}\right)\left(\frac{c}{v_\mathrm{F}}\right)(k_\mathrm{F}d). \tag{2.45}$$

The product of the first three factors is typically on the order of 1, as was seen in Sec. 2.2.1.2. The magnitude of the fourth factor is determined by the density of electrons in the 2D conductor via $k_\mathrm{F} = 4\pi n_0/g$ and distance d between the 2D conductor and the gate. For $n_0 \sim 10^{12}\ \mathrm{cm}^{-2}$ (a typical value for both 2DEG and graphene), $k_\mathrm{F}d = 1$ happens at $d \sim 5\ \mathrm{nm}$. This means that for d much farther than $\sim 5\ \mathrm{nm}$, the restoration force is dominated by Coulomb force (geometric capacitance), while for d much closer than $\sim 5\ \mathrm{nm}$, the restoration force is dominated by electron degeneracy pressure (quantum capacitance). On typical high-mobility 2DEGs, d needs to be farther than $\sim\ 50\ \mathrm{nm}$ due to the semiconductor heterostructure, and it is difficult to obtain such a regime of 2D plasmonic waves propelled dominantly by quantum capacitance. On graphene, $d < 5\ \mathrm{nm}$ may be a possibility, but there has not been a demonstration of such 2D plasmonic wave propagation. In this thesis, we focus on the regime where $k_\mathrm{F}d \gg 1$ so that the 2D plasmonic wave is propelled dominantly by Coulomb force. In this case, we can approximate the gated 2D plasmonic wave dispersion as

$$\omega = \sqrt{\frac{ge^2k_\mathrm{F}d}{4\pi\kappa\epsilon_0\hbar v_\mathrm{F}}}v_\mathrm{F}k_\mathrm{p}. \tag{2.46}$$

If we compare the ungated dispersion relation (Eq. (2.40)) and the gated dispersion relation (Eq. (2.46)) in the regime where Coulomb force dominates, we see that Eq. (2.46) has a much lower slope compared to Eq. (2.40) because we assumed the gate to be 'proximately' gated ($k_\mathrm{p}d \ll 1$). In other words, compared at the same frequency, the gated 2D plasmonic wave exhibits much shorter wavelength (much slower phase velocity) compared to the ungated case because the proximate gate shortens the range of Coulomb interactions in the 2D conductor. Using this principle, 2D plasmonic wave propagation ~ 700 times slower than the speed of light was demonstrated [11], which we will discuss in more detail in Chapter 3.

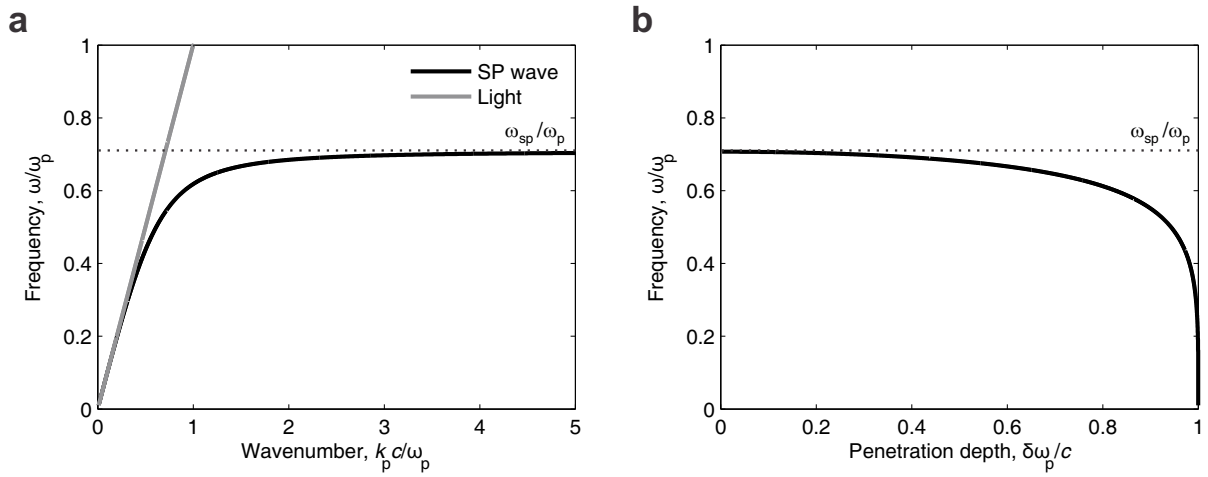

Figure 2.5: Dispersion relation (**a**) and penetration depth (**b**) calculated for surface plasmonic wave at the interface between a hypothetical metal with $\kappa_{\rm m} = 1 - \omega_{\rm p}^2/\omega^2$ and vacuum ($\kappa_{\rm d} = 1$). $\omega_{\rm sp} = \omega_{\rm p}/\sqrt{2}$ in this situation.

2.2.3 Comparison to Surface Plasmonic Waves

Surface plasmonic wave on the surface of 3D metal is a hybridized mode of electromagnetic wave and 3D bulk plasmonic oscillation confined at the surface. Denoting the dielectric constant of the metal as $\kappa_{\rm m}$ and that of the dielectric above the surface as $\kappa_{\rm d}$, one can find the dispersion relation for this wave by solving the Maxwell's equations with appropriate boundary conditions at the surface [6]:

$$k_{\rm p} = \frac{\omega}{c}\sqrt{\frac{\kappa_{\rm m}\kappa_{\rm d}}{\kappa_{\rm m}+\kappa_{\rm d}}}. \tag{2.47}$$

By using the appropriate dielectric constant as a function of frequency for the metal, one can obtain the dispersion relation for the surface plasmons. One commonly used choice for illustrative purposes is the lossless Drude conductivity model, which is also used to explain the 3D bulk plasmonic oscillations, where the dielectric constant of the metal as a function of frequency is given by

$$\kappa_{\rm m} = \kappa_\infty - \frac{\omega_{\rm p}^2}{\omega^2}. \tag{2.48}$$

Here, κ_∞ is the frequency-independent dielectric constant due to the background ion structure, and $\omega_{\rm p}^2 \equiv n_{\rm 3D}e^2/m^*\epsilon_0$ is known as the plasma frequency ($n_{\rm 3D}$: density of electrons in the metal).

The resulting dispersion relation is illustrated in Fig. 2.5a. The dispersion has two asymptotes at low frequencies and at frequencies approaching $\omega_{\rm sp}$, known as the surface plasmon resonance frequency.

At low frequencies, far enough away from the 3D bulk plasmonic resonance, the surface wave almost does not 'see' the effect of the metal and propagate mostly in the dielectric (*i.e.*, very weakly confined to the surface), as seen in the almost light-like behavior in the dispersion. As the frequency rises, the interaction between the electromagnetic wave and the 3D bulk plasmonic oscillation increases, causing a more significant deviation of the dispersion curve from the light line. This culminates at the surface plasmon resonance frequency ω_{sp}, which is the frequency at which the effective dielectric constant seen at the surface becomes zero (*i.e.*, $(\kappa_{\mathrm{m}} + \kappa_{\mathrm{d}})/2 = 0$). In the lossless case assumed here, k_{p} approaches infinity as $\omega \to \omega_{\mathrm{sp}}$, or as the effective dielectric constant approaches 0. However, when loss is included in the model for κ_{m}, the effective dielectric constant at the surface can no longer reach 0, and k_{p} assumes a finite maximum value. In experiments performed with real metals, k_{p}/k_0 is typically limited to less than 10 (k_0: wavenumber of electromagnetic wave in vacuum, *i.e.*, $k_0 = \omega/c$) [6]. At frequencies above ω_{sp}, the effective dielectric constant of the surface becomes positive, meaning that the surface can no longer support a confined wave.

Compared to 2D plasmonic waves (Fig. 2.4), this is qualitatively different in a number of ways. The dispersion relations of 2D plasmonic waves deviate from the light dispersion at all frequencies and does not exhibit an asymptotic behavior to a particular resonance frequency. This means 2D plasmonic wave devices can be created in a wide range of frequencies, as opposed to the surface plasmonic wave devices whose frequency is set by the plasma frequency of the metal in use. In addition, the presence of scattering loss in the conductor does not put a fundamental limit to the maximum attainable k_{p} for 2D plasmonic waves. By carefully engineering the structure, very large values of k_{p}, or in other words, very slow propagation of 2D plasmonic waves is possible, with 2D plasmonic wave propagation ~ 700 times slower than the speed of light experimentally demonstrated [11], as mentioned in the previous section.

As an alternative description of surface plasmonic waves, the transmission line model of Fig. 2.1 can also be used to further contrast surface plasmonic waves and 2D plasmonic waves, as the two main energetic components of the surface plasmonic wave on 3D metal are also the kinetic energy of collectively oscillating electrons and the electric potential energy corresponding to Coulomb restoring force [23].

To describe the behavioral difference between the 2D plasmonic wave and surface plasmonic wave on 3D metal, we evaluate the kinetic inductance associated with the surface plasmonic wave. The collective oscillation of electrons in the surface plasmonic wave occurs within the skin of a 3D bulk metal,

whose frequency-dependent skin depth $\delta = 1/\sqrt{k_{\mathrm{p}}^2 - \kappa_{\mathrm{m}}(\omega/c)^2}$ decreases with increasing frequency (Fig. 2.5b) [23]. The kinetic inductance per unit length of this skin of the 3D metal with thickness δ and width W can be derived by considering the inertial acceleration of electrons in the cross section, similar to the derivation of Section 2.1.1.1, as

$$L_{\mathrm{k,skin}} = \frac{m^*}{n_{\mathrm{3D}} e^2} \frac{1}{W\delta}, \tag{2.49}$$

where n_{3D} is conduction electron density per unit volume (for 3D metal, there is no particular reason to consider massless electrons).

Eq. (2.3) for the 2D plasmonic wave and Eq. (2.49) for the surface plasmonic wave both show that kinetic inductance is inversely proportional to the number of electrons contained in the cross section of the propagating wave ($n_0 W$ for 2D or $n_{\mathrm{3D}} W\delta$ for surface plasmons). This can be understood as follows. When there are fewer electrons in the cross section, they need to accelerate to a proportionally higher maximum velocity to produce the same amplitude of current. The total average kinetic energy then becomes larger, as kinetic energy is proportional to the square of velocity while being linearly proportional to the number of electrons. Since we have fixed the current amplitude, the kinetic inductance then should be larger when there are smaller number of electrons in the cross section, according to Eq. (2.4).

To compare the kinetic inductance for 2D plasmonic wave and surface plasmonic wave, we can rewrite Eqs. (2.3) and (2.49) using $k_{\mathrm{F}}^2 = 4\pi n_0/g$ (2D), $k_{\mathrm{F}}^3 = 6\pi^2 n_{\mathrm{3D}}/g$ (3D), and $E_{\mathrm{F}} = \hbar^2 k_{\mathrm{F}}^2/2m^*$, as products of quantities sharing equal dimensions[2]:

$$\begin{aligned} L_{\mathrm{k,2D}} &= \Big(\frac{2\pi\hbar^2}{ge^2}\Big)\Big(\frac{1}{E_{\mathrm{F}}}\Big)\Big(\frac{1}{W}\Big), \\ L_{\mathrm{k,skin}} &= \Big(\frac{3\pi^2\hbar^2}{ge^2}\Big)\Big(\frac{1}{E_{\mathrm{F}}}\Big)\Big(\frac{1}{W}\Big)\Big(\frac{1}{k_{\mathrm{F}}\delta}\Big). \end{aligned} \tag{2.50}$$

When the two expressions are juxtaposed, the $1/k_{\mathrm{F}}\delta$ factor in $L_{\mathrm{k,skin}}$ makes an apparent difference. Since penetration depth δ decreases with frequency, $L_{\mathrm{k,skin}}$ increases with frequency; this is essentially because the reduced δ decreases the number of conduction electrons participating in the surface plasmonic wave.

[2]We here use $L_{\mathrm{k,2D}}$ to denote the 2D kinetic inductance instead of L_{k} that we have used so far. L_{k} and $L_{\mathrm{k,2D}}$ will be used interchangeably when a comparison to other types of kinetic inductance is needed.

In order to fully describe the surface plasmonic wave in the transmission line model of Fig. 2.1, we also need to take the contribution from magnetic inductance into account, which we have so far ignored because the magnetic inductance is typically orders of magnitude smaller than the kinetic inductance for the 2D plasmonic wave. The effect of magnetic inductance can be modeled as a per-unit-length self-inductance $L_{\mathrm{m}} = \mu_0/2k_{\mathrm{p}}W$ [23] in series with the kinetic inductor in Fig. 2.1.

At frequencies below the optics regime (as ω_{p} for most metals lies in the optics regime), δ is large enough to render $L_{\mathrm{k,skin}}$ inappreciable compared to the magnetic inductance L_{m} of the surface plasmonic medium. Hence, it is difficult to observe surface plasmons below the optics regime with 3D metals, and the surface plasmonic dispersion resembles the light dispersion because the dispersion obtained by the ungated C_{c} (Eq. (2.29)) and L_{m} is identical to the light dispersion $\omega = k_{\mathrm{p}}/\sqrt{L_{\mathrm{m}}C_{\mathrm{c}}} = k_{\mathrm{p}}/\sqrt{\kappa\epsilon_0\mu_0}$. The surface plasmonic dispersion curve deviates away from the light dispersion line only towards the optics regime when $L_{\mathrm{k,skin}}$ becomes large enough to dominate L_{m} (Fig. 2.5). By contrast, $L_{\mathrm{k,2D}}$ of the 2D plasmonic medium has no frequency dependency, as there is no such frequency-dependent penetration depth where electrons are perfectly confined into two dimensions. Moreover, $L_{\mathrm{k,2D}}$ is typically many orders of magnitude larger than the magnetic inductance of the 2D conductor. Therefore, the 2D plasmonic wave emerges far below the optics regime, with its dispersion curve deviating significantly away from the light line even at these low frequencies (Fig. 2.4).

Even when $L_{\mathrm{k,skin}}$ becomes appreciable in the optics regime with small enough δ and with more pronounced surface plasmonic dynamics, $k_{\mathrm{F}}\delta$ is still much larger than 1, leaving $L_{\mathrm{k,skin}} \ll L_{\mathrm{k,2D}}$ (in principle, $L_{\mathrm{k,skin}}$ can grow indefinitely as the frequency grows towards ω_{sp} with δ approaching 0, but in practice loss obscures such excitations). The $L_{\mathrm{k,skin}} \ll L_{\mathrm{k,2D}}$ inequality is further enhanced by the fact that E_{F} appearing in both $L_{\mathrm{k,skin}}$ and $L_{\mathrm{k,2D}}$ in Eq. (2.50) is typically much larger with the 3D bulk metal such as gold than with semiconductor 2DEG or graphene. Overall, $L_{\mathrm{k,skin}}$ even in the optics regime is typically far smaller than $L_{\mathrm{k,2D}}$ by $2 \sim 3$ orders of magnitude. This translates to very slow propagation of 2D plasmonic waves as mentioned earlier, as $v_{\mathrm{p}} \propto 1/\sqrt{L_{\mathrm{k,2D}}}$.

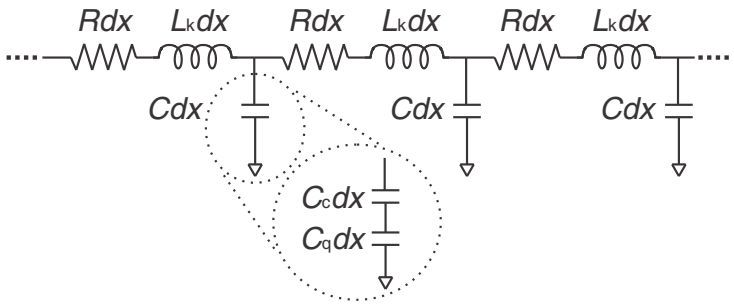

Figure 2.6: Lossy transmission line model of a 2D plasmonic medium [16]. If the medium is gated, the gate serves as the ground. For an ungated medium, the ground is the potential of the free space far away from the 2D medium. dx: infinitesimal segment length of the 2D plasmonic medium.

2.3 Temperature and Loss Considerations

So far, we have considered the behavior of 2D plasmonic wave at $T = 0$ K and without loss. In the absence of loss, the wave propagates indefinitely with a purely imaginary propagation constant (Eq. (2.35)). As sources of loss are introduced into the system via impurities, defects, phonons at finite temperatures, etc.[3], electron scatterings cause the plasmonic wave to be damped, or in other words, finite real part is introduced into the propagation constant. In this section, we discuss how these different types of loss can be integrated into the transmission line model for 2D plasmonic waves, how the loss mechanisms vary with temperature and how that affects the usefulness of 2DEG or graphene as a GHz 2D plasmonic medium, and how the temperature affects other parameters in the model.

2.3.1 Loss Model and Temperature Dependence

Electron scatterings in the 2D plasmonic medium are manifested as a per-unit-length resistance R in electrical measurements[4], which can be added to the transmission line model in series with $L_{\rm k}$ as shown in Fig. 2.6.

Transmission line theory [25] can be used to obtain the generalized complex propagation constant

[3]We here do not consider the loss mechanisms due to Landau damping or interband transitions, as many practical plasmonic applications would be considered at frequencies where such loss mechanisms do not occur.

[4]In the literature, the loss in 2D systems is reported in a number of different forms, such as in sheet resistivity ρ, sheet conductivity σ, mobility μ, or relaxation time τ. They are related to each other via $R = \rho/W = 1/\sigma W$, $\sigma = n_0 e\mu$, and $\tau = (\mu/e)(\hbar k_{\rm F}/v_{\rm F})$, where a n_0-dependent definition of μ was assumed.

and characteristic impedance, given by

$$\gamma = \sqrt{(R + i\omega L_{\mathrm{k}})(i\omega C)}, \tag{2.51}$$

$$Z_0 = \sqrt{\frac{R + i\omega L_{\mathrm{k}}}{i\omega C}}. \tag{2.52}$$

In contrast to the lossless expressions of Eqs. (2.35) and (2.37) that were purely imaginary and real, respectively, these expressions are in general complex. The qualitative behavior of these quantities are strongly dependent on whether the impedance of the resistance is larger than or smaller than the impedance of the kinetic inductance. This is captured by the quality factor Q of the 2D plasmonic wave, which is also obtained from the transmission line theory as

$$Q(\omega) = \frac{\omega L_{\mathrm{k}}}{R} = \omega\tau, \tag{2.53}$$

where τ, the relaxation time due to electron scatterings, factors in through R. As the definition of quality factor is $\omega \times \frac{\text{Energy Stored}}{\text{Power Loss}}$, this signifies how many oscillations the 2D plasmonic wave can make before decaying due to electron scatterings. The plasmonic dynamics can be observed as far as Q is not too far below 1. If $Q \gg 1$, many cycles of collective electron oscillation are sustained between scattering events, making the plasmonic wave very easily observable. If $Q < 1$, the plasmonic behaviors are still observable, but it becomes increasingly difficult as Q becomes smaller.

τ is affected by a number of different scattering mechanisms, each showing a different dependence on the temperature. In particular, the scatterings due to impurities and defects show a relatively weaker dependence on the temperature compared to the scatterings due to acoustic and optical phonons. The effect of these different scattering types on the overall effective relaxation time τ is usually approximated using Matthiessen's rule assuming that each type of scattering is independent from others:

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{impurity}}} + \frac{1}{\tau_{\text{defect}}} + \frac{1}{\tau_{\text{acoustic phonon}}} + \frac{1}{\tau_{\text{optical phonon}}} \cdots . \tag{2.54}$$

This means the overall τ is dominated by the contribution from the scattering mechanism with the shortest relaxation time.

Here, we briefly discuss the temperature dependence of these scattering mechanisms for high-mobility

GaAs/AlGaAs 2DEG (to appear in Chapter 3) and graphene (to appear in Chapter 4), in relation to the practically useful temperature and frequency regime set by these temperature dependencies. The scattering mechanisms in high-mobility GaAs/AlGaAs 2DEG are dominated by extrinsic effects such as impurities, defects, etc. at very low temperatures (liquid helium temperatures or lower), and by longitudinal optical phonons at higher temperatures ($\sim$ 100 K and higher) [1]. The GaAs/AlGaAs 2DEG samples used in this thesis had $\tau \sim 150$ ps at cryogenic temperatures, meaning that $Q > 1$ at ~ 1 GHz or higher. At elevated temperatures, the decrease in τ due to the optical phonons as a function of temperature is very rapid, rendering the room temperature τ to be about 0.08 ps, 3 orders of magnitude smaller than the value at cryogenic temperatures. This means that the GaAs/AlGaAs 2DEG is inappropriate as a GHz 2D plasmonic medium at room temperatures.

In contrast, such adverse effect of optical phonons is absent in graphene, and the main contributors to scattering are longitudinal acoustic phonons and extrinsic scatterers [29]. Scatterings due to the acoustic phonons scale almost linearly, leading to around $\rho = 50\ \Omega/\square$ at room temperatures. In most cases except the cleanest mechanically exfoliated graphene on special substrates (or suspended), the scattering is dominated by extrinsic scatterers, and the overall τ shows a very weak dependence on the temperature, varying by only $2 \sim 3$ times between cryogenic and room temperatures. Even with the cleanest samples, the acoustic phonon scattering may dominate at room temperatures, but the transition from phonon-limited regime to extrinsic-scatterer-limited regime does not happen at very low temperatures. The graphene sample used in this thesis had $\tau \sim 5$ ps at cryogenic temperatures, which decreased by about 4 times at room temperatures. This means that the graphene device just barely allows us to observe plasmonic behaviors at GHz frequencies.

2.3.2 Temperature Dependence of L_k and C_q

The kinetic inductance L_k and quantum capacitance C_q introduced in the earlier sections were both calculated at $T = 0$ K. As will be seen shortly, Eqs. (2.28) and (2.33) calculated at $T = 0$ K suffice for the purposes of this thesis, but it is worthwhile discussing the variations of L_k and C_q at elevated temperatures.

L_k at an arbitrary temperature can be obtained by considering the semiclassical treatment of intraband

electrical conductivity [30],

$$\sigma_{xx} = e^2 \int \frac{d^d\mathbf{k}}{(2\pi)^d} g \frac{v_{\mathbf{k},x}^2 \left(-\frac{\partial f(\varepsilon_\mathbf{k}-\mu)}{\partial \varepsilon_\mathbf{k}}\right)}{\tau^{-1} + i\omega}, \tag{2.55}$$

where d is the number of dimensions, and $f(\varepsilon_\mathbf{k} - \mu) = 1/(e^{(\varepsilon_\mathbf{k}-\mu)/k_\mathrm{B}T} + 1)$ is the Fermi-Dirac distribution with chemical potential μ. L_k can be found by noting that in the limit of $\tau \to \infty$ (*i.e.*, no collisions), $\sigma_{xx} \to (i\omega L_\mathrm{k} W)^{-1}$. C_q can be obtained by first calculating n_0 at an arbitrary temperature as

$$n_0 = \int \frac{d^d\mathbf{k}}{(2\pi)^d} g f(\varepsilon_\mathbf{k} - \mu), \tag{2.56}$$

and noting that $C_q = e^2 \frac{\partial n_0}{\partial \mu} W$. In order to facilitate the calculation, we introduce normalized parameters $\xi = \varepsilon/k_\mathrm{B}T$ and $\eta = \mu/k_\mathrm{B}T$, with which $f(\xi - \eta) = 1/(e^{\xi-\eta} + 1)$ and $\frac{\partial f(\xi-\eta)}{\partial \eta} = -\frac{\partial f(\xi-\eta)}{\partial \xi}$, and use the identity $\int_0^\infty d\xi \xi^s f(\xi - \eta) = -\Gamma(1+s)\mathrm{Li}_{1+s}(-e^\eta)$, where $\Gamma(z)$ is the gamma function and $\mathrm{Li}_n(z) = \sum_{k=1}^\infty \frac{z^k}{k^n}$ is the polylogarithm function that satisfies $\frac{d}{dx}\mathrm{Li}_n(x) = \frac{1}{x}\mathrm{Li}_{n-1}(x)$. For a massive 2D electron gas with $\varepsilon = \frac{\hbar^2 k^2}{2m^*}$, it is then a straightforward exercise to show that

$$n_0 = \frac{g m^* k_\mathrm{B} T}{2\pi\hbar^2} \left[-\mathrm{Li}_1(-e^{\mu/k_\mathrm{B}T})\right], \tag{2.57}$$

$$L_\mathrm{k} = \frac{2\pi\hbar^2}{g e^2 k_\mathrm{B} T} \frac{1}{-\mathrm{Li}_1(-e^{\mu/k_\mathrm{B}T})} \frac{1}{W}, \tag{2.58}$$

$$C_\mathrm{q} = \frac{g m^* e^2}{2\pi\hbar^2} \left[-\mathrm{Li}_0(-e^{\mu/k_\mathrm{B}T})\right] W. \tag{2.59}$$

We note that Eqs. (2.57) and (2.58) satisfy $L_\mathrm{k} = \frac{m^*}{n_0 e^2} \frac{1}{W}$ (identical to Eq. (2.3) calculated at $T = 0$ K) at all temperatures.

For graphene, we need to consider the contributions from the conduction band (electrons) and the valence band (holes) separately, as electrons and holes may exist simultaneously at an elevated temperature. For holes, we may use the same $\int_0^\infty d\xi$ as the integration limits but take $f(\xi + \eta)$ as the distribution function, with which $-\frac{\partial f(\xi+\eta)}{\partial \eta} = -\frac{\partial f(\xi+\eta)}{\partial \xi}$. We also note that while both electrons and holes contribute positively to the conductivity in Eq. (2.55), their charges cancel each other to reduce the effective density (this may be seen as the total charge density in an electrostatically gated graphene, which is a fixed

quantity in such a situation) in Eq. (2.56). Noting these two, n_0, L_k [31], and C_q [19] are evaluated to be

$$n_0 = \frac{g(k_\mathrm{B}T)^2}{2\pi(\hbar v_\mathrm{F})^2}\left[-\mathrm{Li}_2(-e^{\mu/k_\mathrm{B}T}) + \mathrm{Li}_2(-e^{-\mu/k_\mathrm{B}T})\right], \tag{2.60}$$

$$L_\mathrm{k} = \frac{4\pi\hbar^2}{ge^2k_\mathrm{B}T}\frac{1}{-\mathrm{Li}_1(-e^{\mu/k_\mathrm{B}T}) - \mathrm{Li}_1(-e^{-\mu/k_\mathrm{B}T})}\frac{1}{W}, \tag{2.61}$$

$$C_\mathrm{q} = \frac{ge^2k_\mathrm{B}T}{2\pi(\hbar v_\mathrm{F})^2}\left[-\mathrm{Li}_1(-e^{\mu/k_\mathrm{B}T}) - \mathrm{Li}_1(-e^{-\mu/k_\mathrm{B}T})\right]W. \tag{2.62}$$

Interestingly, $L_\mathrm{k}C_\mathrm{q} = 2/v_\mathrm{F}^2$ holds at all temperatures for graphene, while it holds only at $T = 0K$ for a massive electron gas.

In order to see the temperature dependence of these expressions, Figs. 2.7 and 2.8 display the ratio of L_k expressions and C_q expressions at finite and zero temperatures, showing that the expressions at zero temperature are almost identical to the ones at finite temperature as long as $E_\mathrm{F} > 5k_\mathrm{B}T$. As practical values of E_F are on the order of 0.01 eV for high-mobility GaAs/AlGaAs 2DEG and on the order of 0.1 eV for graphene, $E_\mathrm{F} > 5k_\mathrm{B}T$ condition is satisfied for both cases at cryogenic (liquid helium) temperatures. At room temperature, as $k_\mathrm{B}T = 0.026$ eV, the zero temperature expression for C_q cannot be used for the GaAs/AlGaAs 2DEG, and there may be notable corrections to the L_k and C_q expressions for graphene as $E_\mathrm{F} \approx 5k_\mathrm{B}T$. However, due to the rapid decrease of τ at elevated temperatures in GaAs/AlGaAs 2DEG, we limit ourselves to cryogenic temperatures for the devices made with GaAs/AlGaAs 2DEG, while with graphene, one may be able to observe these temperature-dependent effects.

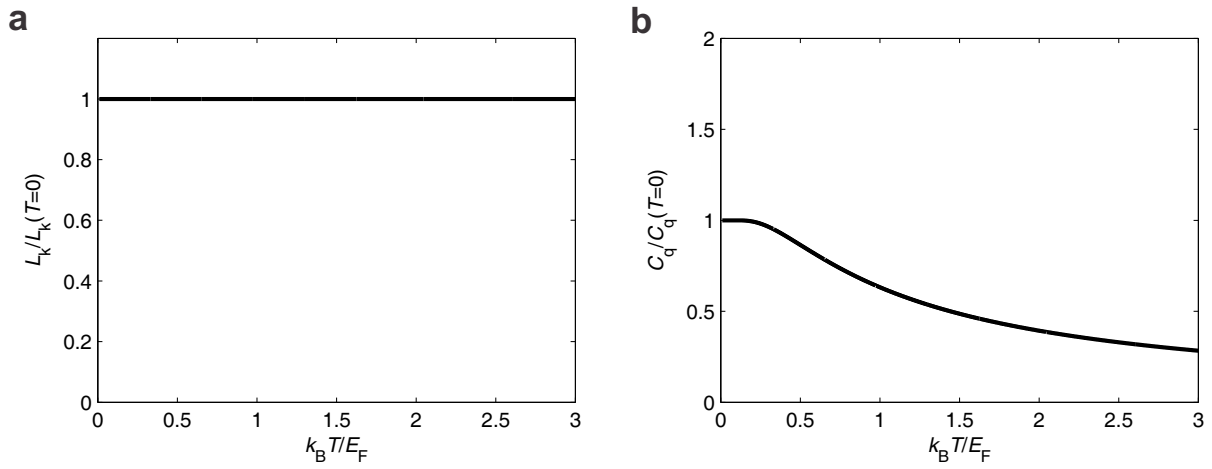

Figure 2.7: Ratios of finite temperature expressions and zero temperature expressions for $L_{\rm k}$ (**a**) and $C_{\rm q}$ (**b**) in a massive 2D electron gas. Constant n_0 is assumed as temperature is varied; $E_{\rm F} = \frac{\hbar^2}{2m^*}\frac{4\pi n_0}{g}$ is a fixed quantity while μ varies.

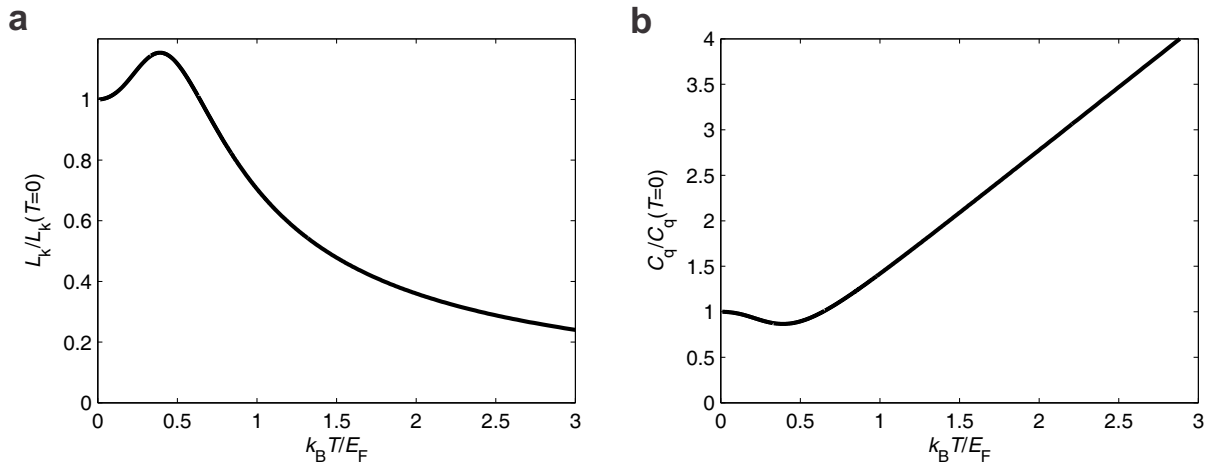

Figure 2.8: Ratios of finite temperature expressions and zero temperature expressions for $L_{\rm k}$ (**a**) and $C_{\rm q}$ (**b**) in graphene. Constant n_0 is assumed as temperature is varied; $E_{\rm F} = \hbar v_{\rm F}\sqrt{4\pi n_0/g}$ is a fixed quantity while μ varies.

Chapter 3

2D Plasmonics in Semiconductor 2DEG

A 2D plasmonic medium can be readily shaped into a designed planar geometry by using standard fabrication technology. 2D plasmonic waves can then be manipulated by reflections, interferences, and superposition according to the geometry to create various 2D plasmonic circuits and metamaterials [11, 13, 14]. In this chapter, we discuss the design, realization, and measurements of a number of ultra-subwavelength 2D plasmonic devices created using GaAs/AlGaAs 2DEG, including bandgap crystals analogous to photonic crystals in optics, interferometers, and negatively refracting metamaterials with extremely large negative index of refraction.

3.1 2D Plasmonic Crystals and Interferometers

2D plasmonic circuits[1] in the electronics frequency regime (GHz) can bring the extreme subwavelength confinement advantages of 2D plasmonic waves to electronics in direct interface with them. To this end, we first construct 2D plasmonic waveguides consisting of a highly-degenerate GaAs/AlGaAs 2DEG at $T = 3$ K as a 2D plasmonic medium and a metallic top gate 80 nm above the 2DEG as a 'plasmonic ground,' thus creating localized plasmonic input and output terminals (Fig. 3.1). Due to these local terminals, the plasmonic waveguides are directly and non-resonantly excited over a continuous frequency range, driven by the voltage and current of electromagnetic signals guided to a local terminal. This direct local coupling of electromagnetic signals to the full spectrum of plasmons allows complete free-

[1] Large portions of this section are derived from a paper in publication by W. F. Andress, *et al.* [11]. The author has participated in large part to the theory, simulation, and experiment of this work.

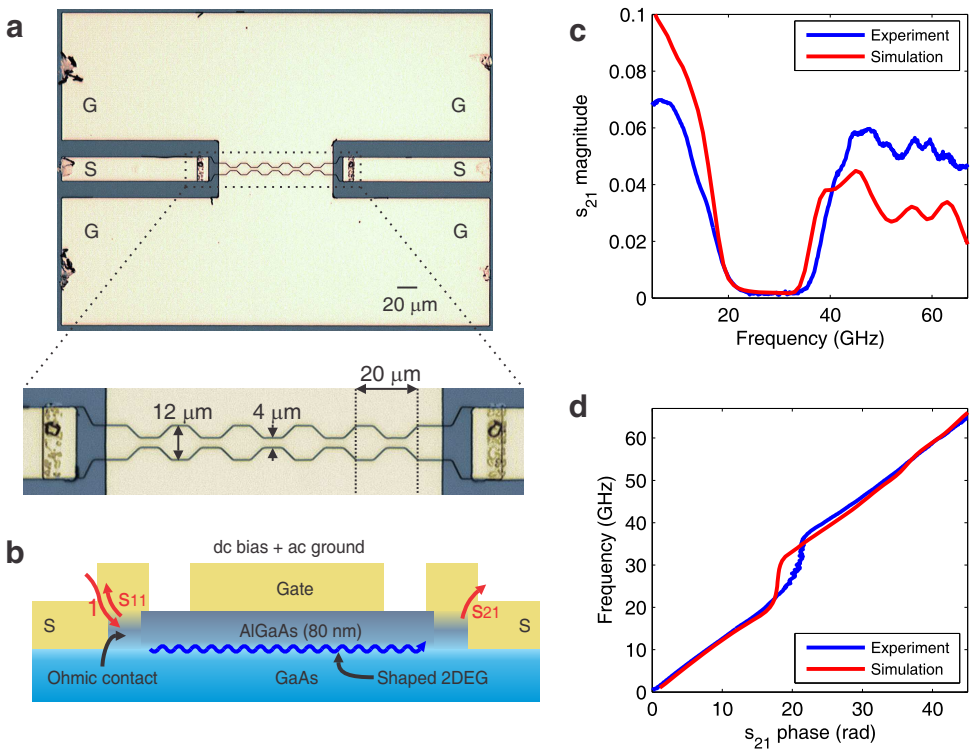

Figure 3.1: 2D plasmonic crystal with one-dimensional periodicity [11]. **a**, Optical image of plasmonic crystal structure. Coplanar waveguides (CPWs; signal line 'S' and ground lines 'G') that direct electromagnetic waves to and from the device are also shown. **b**, Diagram showing the cross section of the device. Electromagnetic waves (red curves) couple to plasmonic waves (blue curve), which propagate underneath the gate through the periodically shaped 2DEG. **c**, Measured (blue) and simulated (red) transmission (s_{21}) magnitude through the device at 0.5 V bias, showing plasmonic bandgap from 24 to 34 GHz. **d**, Measured (blue) and simulated (red) s_{21} phase curve (plasmonic dispersion diagram).

dom to route and manipulate plasmons in between these local terminals. This manipulation is done by freely shaping the planar geometry of 2DEG, the plasmonic signal medium itself, within the plasmonic waveguide. Plasmons are elaborately routed according to the designer geometry to undergo specific reflections and interferences , leading to GHz plasmonic circuits with a range of functionalities, such as plasmonic-electromagnetic interferometers, plasmonic cavity, and plasmonic bandgap crystals with designer bandgap characteristics and with one- and two-dimensional periodicities. These plasmonic circuits achieve extreme subwavelength confinement due to the large 2D kinetic inductance and the top gate, or plasmonic ground, which shortens the Coulomb interaction range of electrons, occupying areas up to 440,000 times smaller than their electromagnetic counterparts, which are ubiquitous in electronics but not subject to Moore's-law scaling enjoyed by transistors.

In operating principle and functional versatility, our passive 2D plasmonic circuits differ from the

prior passive 2D plasmonic devices [20, 32–36], which are mostly used as resonant detectors from GHz to THz to far infrared. They do not utilize the geometric shaping of the plasmonic medium to elaborately route plasmons in deriving their functionalities; they rather employ an overall exposure of the 2D plasma to electromagnetic waves, with the absorption occurring at some discrete set of resonant coupling frequencies. This resonant coupling arises from the periodicity of grating couplers above the plasma [32, 35, 36], or from the plasma's spatial confinement [20, 33, 34]; one example of the latter type involves GHz electromagnetic signals delivered to a gate metal overlying a bounded plasma, maximally absorbed at the plasma resonances [20]. A different, active, type of 2D plasmonic devices attain THz emission via instability-mediated self-sustained 2D plasma oscillation [37, 38], or use its inverse mechanisms to perform THz detection [39]; none of them utilizes the shaping of the plasmonic medium, thus, it would be an interesting possibility to combine our passive shaping principle with the active mechanisms in the instability regime.

3.1.1 2D Plasmonic Crystals with One-Dimensional Periodicity

A starting example of our approach is the plasmonic bandgap crystal (Fig. 3.1a). Here the plasmonic waveguide consists of the 2DEG as a plasmonic signal medium and a top gate metal 80-nm above (Figs. 3.1a-b), where the gate will act as a 'plasmonic ground' if configured properly. The planar geometry of the 2DEG is physically shaped by mesa etching into a channel of periodically varying width, while the gate is not patterned. The plasmonic waveguide is excited at either of the local terminals on the left and right, each of which connects the shaped 2DEG to the signal line (labeled 'S', Figs. 3.1a-b) of a purely electromagnetic on-chip metallic coplanar waveguide (CPW), via a Ni/Au/Ge ohmic contact. Each CPW, deposited over the etched insulating region, consists of the signal line and two ground lines (labeled 'G', Fig. 3.1a). The CPWs' ground lines connect with the gate, allowing this gate to act as a plasmonic ground. This ground sharing prevents the gate itself from becoming a signal path, such that a purely plasmonic waveguide continues from the purely electromagnetic CPWs (this will be discussed in more detail in Sec. 3.1.3); this is an important principle to make our plasmonic circuits work, and is different from works that use a gate as an overall excitation source of the underlying plasma [20].

When the plasmonic crystal (Fig. 3.1a) is excited at the left terminal, a plasmonic wave is launched rightward. Transmission through the crystal is measured as the microwave scattering parameter s_{21} at

the right terminal up to 66 GHz. Due to the crystal periodicity, a bandgap ($24 \sim 34$ GHz) occurs in the measurement (Fig. 3.1c). This bandgap arises from the first Brillouin-zone boundary, where the 20-μm period equals half the plasmonic wavelength, $\lambda_{\mathrm{p}}/2$. Since the frequency corresponding to the zone boundary must lie in this band, the plasmonic velocity v_{p} lies somewhere between $c/220$ and $c/310$, thus, λ_{p} is 265 ± 45 times smaller than the free-space electromagnetic wavelength, λ_0. The direct, non-resonant, interface of plasmonic circuits with electronics (the measurement setup here) also makes the direct measurement of the dispersion relation possible, which is simply s_{21}'s phase that is proportional to Bloch wavenumber k_{p} measured over a continuous frequency range. As expected, the plasmonic crystals' dispersion (Fig. 3.1d) contains a discontinuity over the bandgap. The spurious phase deviation within the bandgap arises from parasitic signals. From the dispersion curve's passband slope, we extract $v_{\mathrm{p}} \sim c/(300 \pm 30)$, with precision limited by the frequency-dependent effects of the ohmic contacts on s_{21}'s phase (v_{p} will be extracted more accurately later). The next bandgap does not occur until the third Brillouin-zone boundary [40], beyond our measurement range. Simulated s-parameters obtained using Sonnet frequency-domain field solver with estimated values of kinetic inductance and ohmic contact parameters resemble the measured s-parameters (Figs. 3.1c-d) in the overall bandgap character, with very similar bandgap positions; the differing transmission amplitudes are due to the estimated nature of the ohmic contact parameters used in the simulation; the differing s_{21} phases (Fig. 3.1d, x-axis) within the bandgap are not surprising, as they in general spuriously arise when the signal transmission is low in the bandgap.

The plasmonic crystal serves to introduce a general geometric-shaping principle, which allows a great deal of versatility. For instance one can begin to add subtle variation to the crystal's geometry to produce a conspicuous behavioral difference. Fig. 3.2a shows a variation on the plasmonic crystal, where the transitions between the thin and thick sections are made abruptly. The plasmonic dynamics here cannot be simply viewed as a perturbed horizontal routing as in Fig. 3.1; one must also consider the vertical routing of plasmons to and from the ends of the thick sections, or vertical stubs (each thick section contains two symmetric stubs). These vertical stubs act as plasmonic cavities, which resonate when plasmons routed toward and reflected from the stub ends interfere to form standing waves. The repetition of the plasmonic cavity is expected to result in another bandgap at around the cavity resonance frequency. The measured transmission (Figs. 3.2b-c) indeed includes not only the Brillouin-zone-boundary bandgap

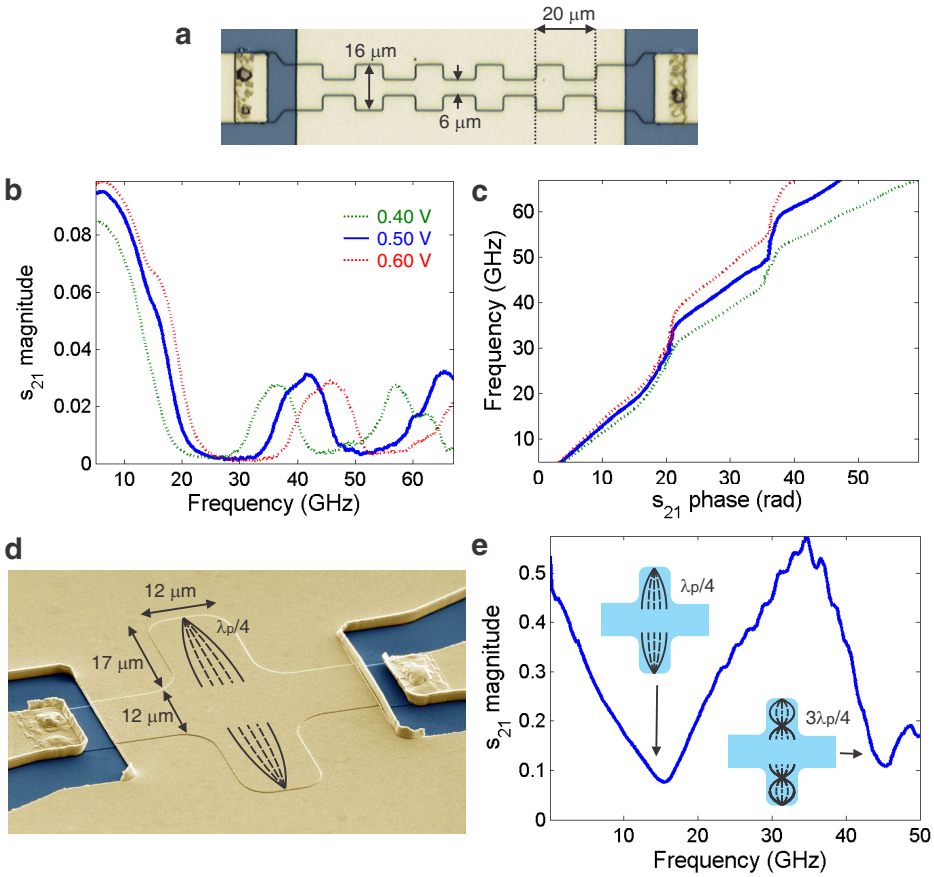

Figure 3.2: 2D plasmonic cavity array [11]. **a**, Optical image of a periodic array of cavities. **b**, s_{21} magnitude curves of the device in **a** at different bias voltages. **c**, s_{21} phase curves (plasmon dispersion diagrams) of the device in **a**. Different slopes at each bias voltage reflect different plasma wave velocities. **d**, Scanning electron micrograph of a stand-alone plasmonic cavity circuit (false-colored). The current amplitude profile of a $\lambda_{\mathrm{p}}/4$ standing-wave resonance in each vertical stub is illustrated. **e**, Measured voltage transmission (s_{21}) magnitude through the device at 0.5 V bias, with the effects of ohmic contacts de-embedded. The two resonance modes are illustrated.

as in Fig. 3.1c, but also the additional bandgap, centered about 52 GHz, from the cavity resonance.

To verify this cavity behavior, we build a stand-alone cavity structure with two vertical stubs (Fig. 3.2d). At the frequency where the stub-lengths are $\lambda_{\rm p}/4$, a standing-wave resonance should emerge. In this situation, the standing wave's current [voltage] amplitude is zero [maximal] at the stub ends (which act as open circuits; fringing capacitances to the gate are negligible) and maximal [zero] at the stub inlets (Fig. 3.2d). Thus at the $\lambda_{\rm p}/4$ frequency, the impedance at the inlets becomes zero, and the entire wave is absorbed into the stubs, with nothing transmitted to the right terminal. The s_{21} measured from the stand-alone cavity structure (Fig. 3.2e) indeed shows this characteristic, with the $\lambda_{\rm p}/4$ absorption at 15 GHz. Given the 17-μm stub length, this corresponds to $\lambda_{\rm p} = 68\ \mu{\rm m} = \lambda_0/294$, thus, $v_{\rm p} = c/294$. Another standing-wave resonance expected when the stub-length is $3\lambda_{\rm p}/4$ indeed appears at 45 GHz. The cavity may be viewed as a 'plasmon trap' with frequency-dependent potential in the particle picture. This resonance from the cross-shaping is independent of the input coupling. The more precise estimate of $v_{\rm p} = c/294$ obtained from the cavity behavior is within the range deduced from the bandgap of Fig. 3.1c and also within that deduced from the passband slope of Fig. 3.1d. Also we can now verify that the additional bandgap (Fig. 3.2b) from the crystal of Fig. 3.2a originates from its vertical cavities; if we continue to assume $v_{\rm p} = c/294$ (although different structures may support slightly different $v_{\rm p}$ since mesa boundaries influence electron density), the $\lambda_{\rm p}/4$ resonance in the 5-μm long cavities should occur at ~ 51 GHz, which indeed falls within the second bandgap (Fig. 3.2b).

The extremely slow plasmonic velocity $v_{\rm p} = c/294$, and thus the extremely short plasmonic wavelength $\lambda_{\rm p} = \lambda_0/294$, drastically miniaturizes the 2D plasmonic circuits relative to their electromagnetic counterparts. As mentioned earlier, this extreme subwavelength confinement is attained by the inherently large 2D kinetic inductance of the plasmonic medium and the top gate shortening the Coulomb interaction range of electrons in the 2DEG. These two factors in conjunction result in $v_{\rm p} \sim 4v_{\rm F}$, for which the Coulomb restoring force is ~ 30 times stronger than the restoring force due to degeneracy pressure (*i.e.*, $C_{\rm c} \sim C_{\rm q}/30$). Using the gate as a dc bias, we can alter the 2DEG's electron density, thus, kinetic inductance by changing the gate's bias relative to the signal line, tuning $v_{\rm p}$ (and simultaneously, $v_{\rm F}$, keeping $v_{\rm p} \sim 4v_{\rm F}$) from $c/187$ at 0.8 V to $c/660$ at -0.13 V. This tunability can be seen in Figs. 3.2b-c, where the bandgap frequencies and dispersion slopes are altered with bias. With $\lambda_{\rm p}$ as small as $\lambda_0/660$, plasmonic circuits can occupy up to ~440,000 times less area than their electromagnetic counterparts.

Another advantage is apparent from the quality factor, Q, at the $\lambda_{\mathrm{p}}/4$ and $3\lambda_{\mathrm{p}}/4$ resonances, estimated as 4.1 and 13 from their full-width half-maximums (Fig. 3.2e). The almost linear increase of Q with frequency is characteristic of plasmons, for which $Q/2\pi$ equals the frequency times momentum relaxation time τ (Eq. (2.53)), indicating how many times electrons oscillate between collisions with crystal impurities, defects, and phonons. Thus while this work demonstrates prototype plasmonic circuits at 3 K to attain a long enough τ to achieve $Q \sim 10$ at GHz frequencies, their linearly increasing Q suggests an intrinsic scalability of our plasmonic circuits into higher frequencies based on the geometric shaping of the plasmonic medium and the resultant plasmon routing-manipulation, with possible applicability toward the THz gap [41], where room temperature implementation may be feasible. With the recent observation of light-plasmon coupling in graphene at THz and room temperature [12], graphene may offer an especially interesting 2DEG platform on which to apply our shaping-based plasmon manipulation principle.

3.1.2 2D Plasmonic Crystals with Two-Dimensional Periodicity

To illustrate our principle further, we create plasmonic crystals with two-dimensional periodicity, by etching lattices of holes into the 2DEG. Two rectangular lattices and a hexagonal lattice are shown in Fig. 3.3a. The two rectangular lattices have identical unit cells, but the orientation of each lattice relative to the propagation direction of the plasmonic waves differs by $90\ ^{\circ}$. For plasmons traveling rightward in each lattice, a Brillouin-zone-boundary bandgap occurs when the distance between adjacent vertical crystal planes is $\lambda_{\mathrm{p}}/2$. Since these separations are 16, 14, and 12.1 μm in the three lattices and the bandgap in the first structure occurs about 36 GHz ($v_{\mathrm{p}} \sim c/260$), we expect bandgaps in the other two structures near 41 and 47.6 GHz, which indeed fall within the corresponding measured bandgaps (Fig. 3.3b). The frequency positions of the discontinuities in the measured dispersions follow these ratios as well.

Measurements are performed with five additional 2D plasmonic crystals to further confirm the behavioral dependency on the planar geometry. Fig. 3.4a shows a pair of devices sharing the same square-lattice crystal structure of a 14-μm periodicity, but of different overall lengths and widths. As seen in the measured s_{21} magnitudes (Fig. 3.4b), the bandgap positions, which are governed by the distance between two adjacent vertical crystal planes, are very similar. The difference in the overall amplitude

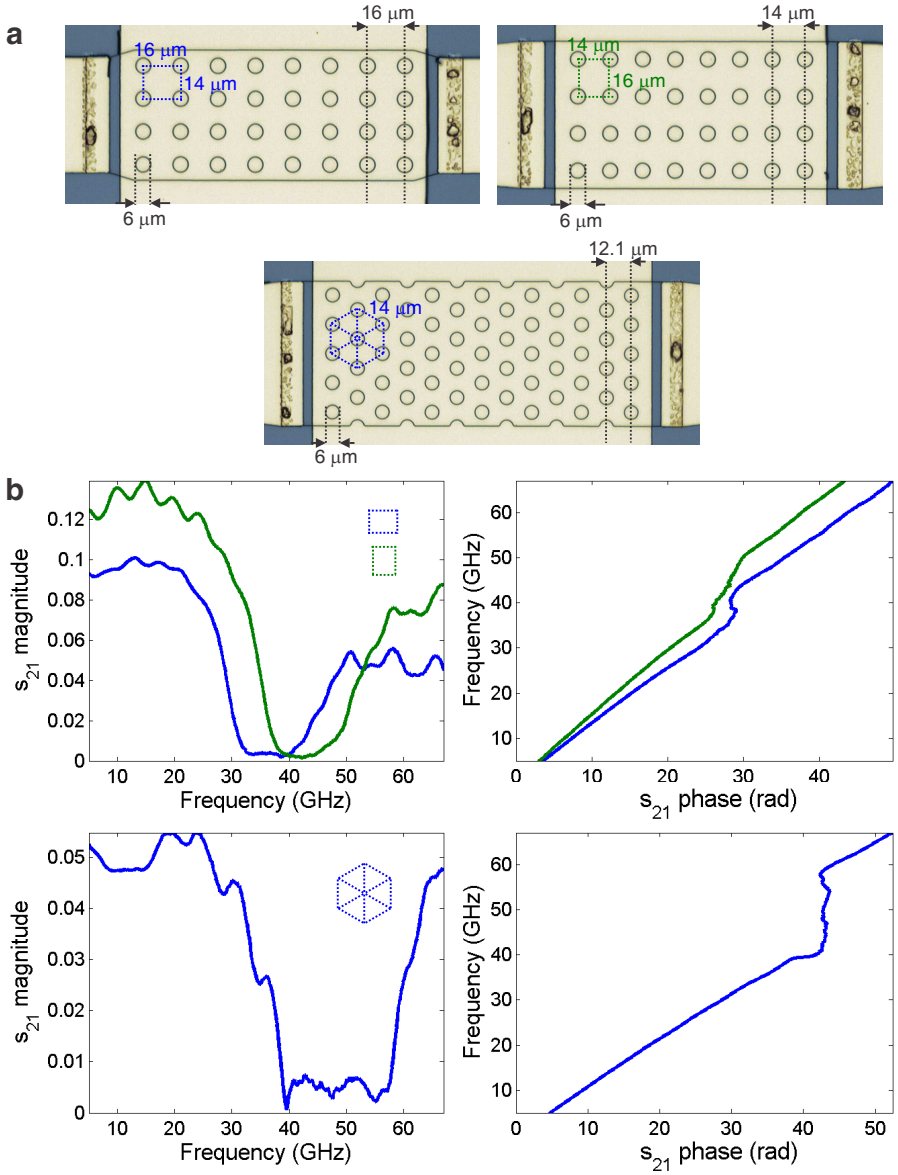

Figure 3.3: 2D plasmonic crystal with two-dimensional periodicity [11]. **a**, Device images of rectangular and hexagonal 2D plasmonic crystals. **b**, s_{21} magnitude (left) and s_{21} phase (right) of rectangular (top) and hexagonal (bottom) crystals at 0.5 V bias. The three devices have similar velocities; note that the phase curve's slope must be multiplied by the device length in order to extract the velocity, hence the difference in slope between the two rectangular orientations despite their similar velocities.

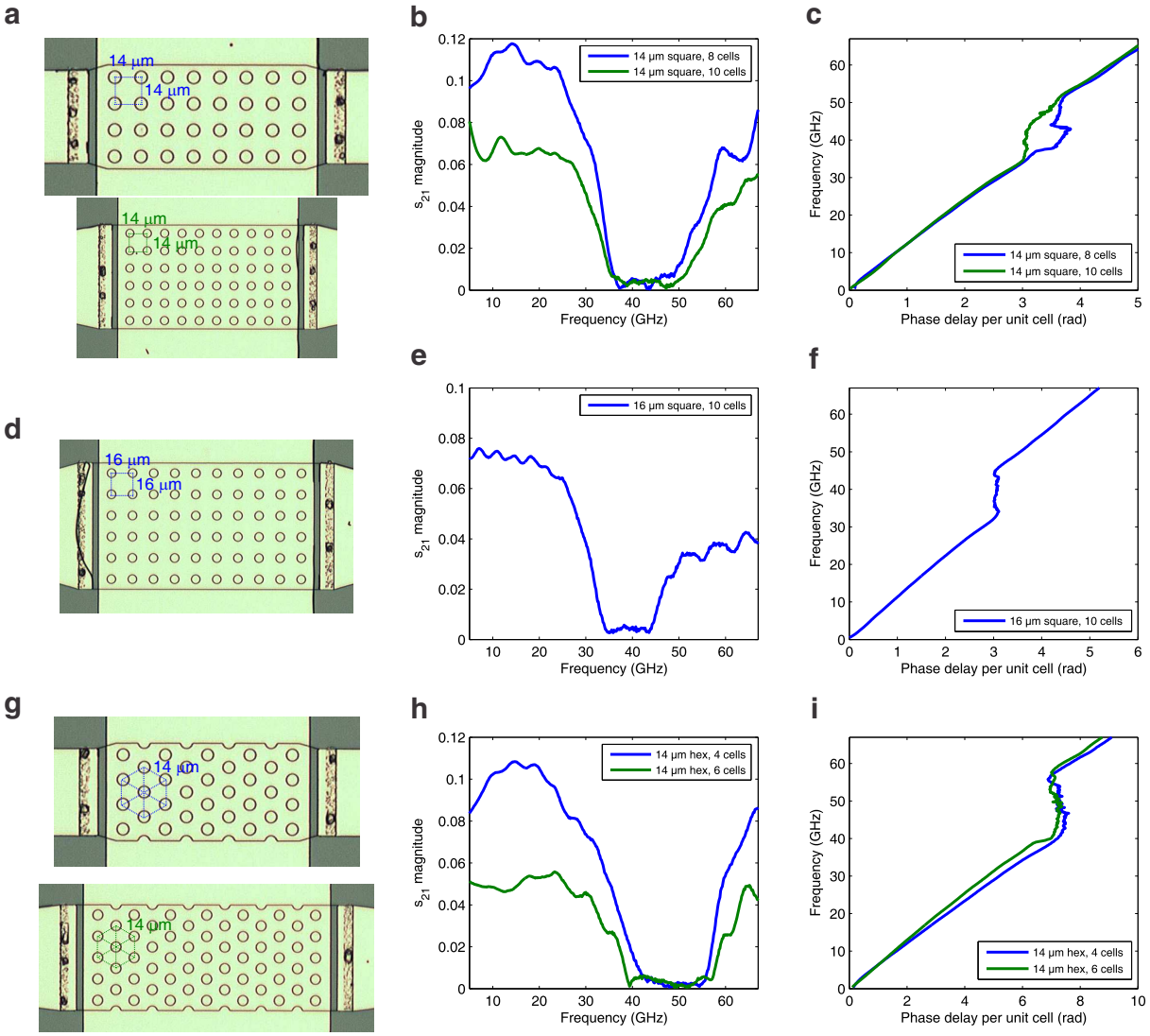

Figure 3.4: 2D plasmonic crystal with two-dimensional periodicity (continued) [11]. **a-c**, Device images, measured s_{21} magnitudes, and s_{21} phase delays per unit cell, respectively, for a pair of devices that share the same square-lattice crystal structure of a 14-μm periodicity (the image scale is different between the two devices), but of different overall lengths and widths. **d-f**, Device image and measurement results for another square-lattice device with a larger periodicity of 16 μm. **g-i**, Device images and measurement results for another pair of devices that share the same hexagonal-lattice crystal structure of a 14-μm periodicity, but of different overall lengths and widths.

is well expected, as one device is longer, thus more lossy, than the other. The measured dispersions (Fig. 3.4c; obtained from the measured s_{21} phase; y-axis is frequency; x-axis is phase delay per unit cell) also indicate very similar bandgap positions (y-axis) and passband dispersions; the differing phase delays per unit cell (x-axis) within the bandgap are not surprising, as they in general spuriously arise from noise and calibration errors when the signal transmission is low. In comparison to these results, the bandgap position of another square-lattice device with a larger periodicity of 16 μm (Fig. 3.4d) is expectedly moved to a lower frequency (Figs. 3.4e,f). Fig. 3.4g shows another pair of devices sharing exactly the same crystal structure (this time, hexagonal lattices with a periodicity of 14 μm), but with different overall lengths and widths. The measured s_{21} magnitudes and dispersions (Figs. 3.4h,i) again show very similar bandgap positions, along with the expected amplitude difference due to different lengths. These new measurements confirm that the key behaviors arise primarily from geometry.

3.1.3 Hybrid Electromagnetic-Plasmonic Interferometer

While all the above circuits exploit interferences of multiple plasma waves, another interesting avenue to explore is interference between plasmonic and purely electromagnetic waves. Fig. 3.5 shows a hybrid electromagnetic-plasmonic interferometer. The 2DEG, excited from the left contact, is split into two smaller 2DEG branches of equal width. The gate that covers the bottom branch is connected to one of the ground lines of the metallic CPW. Therefore, as in all previous circuits of Figs. 3.1 to 3.3, this branch acts as a plasmonic signal path. The top branch, on the other hand, is covered by a gate that is not connected to the CPW ground. Although this gate is biased by a dc probe, at GHz frequencies the inductance of the dc probe and cable prevents any current from flowing through it. Therefore, this gate effectively functions as a 'floating' gate, and becomes a path for propagation of electromagnetic waves, as opposed to plasmonic waves (this will be discussed in more detail below). The signal is therefore divided into a slow plasmonic wave in one path, and a fast electromagnetic wave in the other, with amplitudes of similar magnitude. As the excitation frequency is swept up to 50 GHz, the phase of the slow plasmonic wave rapidly changes while the phase of the fast electromagnetic wave remains nearly the same, resulting in interference. The measured interference pattern (Fig. 3.5b) features a sequence of dips, which occur when the two types of waves interfere destructively, or when the plasmonic wave's phase shift differs by an odd multiple of a half cycle relative to the nearly constant phase of the electromagnetic wave. These

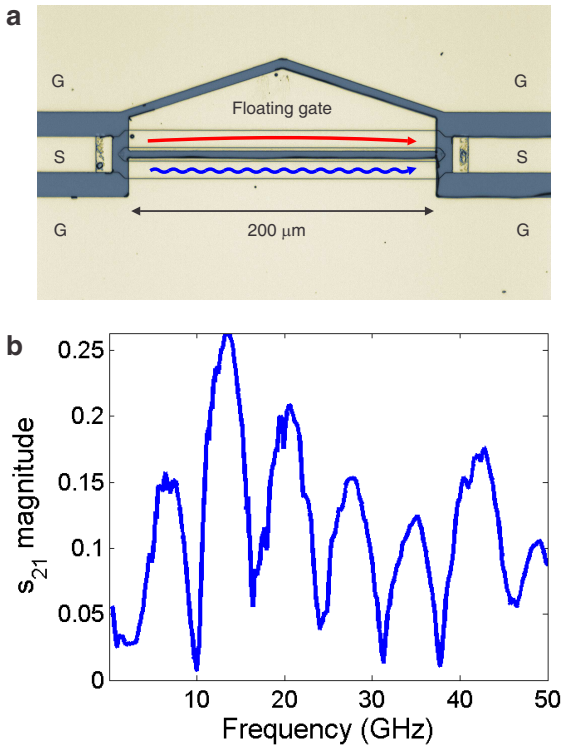

Figure 3.5: Hybrid electromagnetic-plasmonic interferometer [11]. **a**, Optical image. A purely electromagnetic wave (red curve) propagates *along the floating gate* while a purely plasmonic wave (blue curve) propagates along the 2DEG below the grounded gate that is connected to the ground lines of the CPWs. **b**, s_{21} magnitude curve of the device at 0.8 V bias. Repeated peaks and valleys result from the interference between plasmonic and electromagnetic waves.

dips are separated by 6 GHz, implying v_{p} of $c/238$ in this particular structure at 0.8 V gate bias. Note that while a plasmonic Fabry-Perot resonance could manifest similarly to the measured interferometric pattern, dissipations in the ohmic contacts and 2DEG prevent it in this circuit of Fig. 3.5 as well as all other devices found in Figs. 3.1 to 3.3, except for a very weak Fabry-Perot resonance in the short device of Figs. 3.2d-e.

The electromagnetic propagation in the top branch can be understood in more detail as follows. Since the gate over the top branch is floating, we can see this gate in conjunction with the adjacent

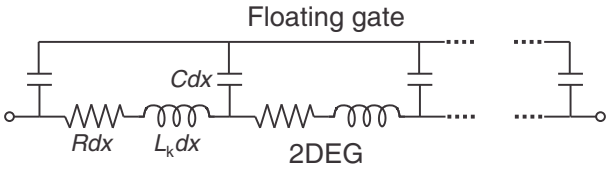

Figure 3.6: Circuit model of the path with floating gate [11].

CPW grounds as forming an electromagnetic transmission line. As the 2DEG below the floating gate is depleted, signals propagating through this 2DEG predominately couple into this fast electromagnetic transmission line. Even if the 2DEG beneath the floating gate is not completely depleted, propagation is still predominantly through the floating gate. To illustrate this more quantitatively, we can create a simple circuit model of this phenomenon. A gated 2DEG can be modeled using its kinetic inductance, scattering resistance, and capacitance to the gate, as in Fig. 3.6. The floating metallic gate may be treated as a short circuit, given its small length relative to the electromagnetic wavelength. The small device length and the absence of ground connection for the floating gate allow us to treat this path simply as a lumped element with an impedance Z *across* it (not to be confused with the local port input impedance). In the regime where $Q > 1$, this impedance can be approximated from Eqs. (2.51) and (2.52) as $Z = 2Z_0 \tanh\big[(\alpha + ik_\mathrm{p})l/2\big]$, where $Z_0 = \sqrt{L_\mathrm{k}/C}$, $\alpha = R/2Z_0$, $k_\mathrm{p} = \omega\sqrt{L_\mathrm{k}C}$, and l is the length of the gate. When the 2DEG is depleted to very low electron density, R and L_k are both increased, and thus the argument of $\tanh$ can become large enough that Z approaches $2Z_0$, which is real like a lumped resistor. This indicates that this branch with the floating gate introduces inappreciable phase delay. It physically corresponds to the excitations beneath one edge of the gate simply traveling into the gate to the other side.

3.1.4 Appendix: Materials and Methods

Our plasmonic circuits are fabricated on GaAs/AlGaAs 2DEG substrates grown by molecular beam epitaxy (MBE). From the 2DEG up, the layer structure is as follows: 48 nm $\mathrm{Al_{0.3}Ga_{0.7}As}$, 26 nm Si-doped $\mathrm{Al_{0.3}Ga_{0.7}As}$, and 6 nm GaAs cap, totaling to 80 nm. At 3.7 K, before processing, the carrier density is $1.54 \times 10^{11}/\mathrm{cm}^2$, with mobility $2.5 \times 10^6 \mathrm{cm}^2/\mathrm{Vs}$ in the dark, and the carrier density is $2.8 \times 10^{11}/\mathrm{cm}^2$ with mobility $3.9 \times 10^6 \mathrm{cm}^2/\mathrm{Vs}$ after illumination.

Mesas defining the shape of a plasmonic circuit are formed by photolithography followed by wet etching by > 80 nm to beneath the 2DEG layer, using 240:8:1 $\mathrm{H_2O : H_2O_2 : H_2SO_4}$. The next photolithography step defines the Ni-Au-Ge ohmic contacts, which overlap the mesas by 6 μm from the ends. After an ammonia dip, the contact metals are deposited by thermal evaporation as follows: 5 nm Ni, 20 nm Au, 25 nm Ge, 10 nm Au, 5 nm Ni, 40 nm Au. After liftoff, the contacts are annealed at $420\,^\circ\mathrm{C}$ for 50 seconds. The final photolithography step defines the coplanar waveguides (CPWs) as well as the

gate that covers the plasmonic medium, which is connected to the ground (G) lines of the CPWs. The signal (S) lines of the CPWs extend just over the ohmic contacts. The gate is separated from the contacts and signal lines by roughly 4-μm wide ungated areas. The CPWs and gate metals, also deposited by thermal evaporation, consist of 8 nm Cr and 500 nm Au.

3.2 2D Plasmonic Negative Index Metamaterials

Negative index metamaterials[2] offer unusual abilities in manipulating electromagnetic waves, such as sub-diffraction-limit focusing [42], light bending in the 'wrong' direction [43], and reversed Doppler and Cerenkov effects [43]. These counterintuitive and technologically gainful behaviors have spurred significant efforts to synthesize a broad array of negative index metamaterials with engineered electric, magnetic, or optical properties [42–51]. Here we demonstrate another route to negative refraction; we exploit the kinetic property, or inertia, of electrons in semiconductor 2DEG, collectively accelerated by electromagnetic waves according to Newton's law of motion, where this acceleration effect manifests as kinetic inductance [20, 52]. Using kinetic inductance to attain negative refraction was theoretically envisioned with three-dimensional metallic nanoparticles [53] and experimentally glimpsed with surface plasmons on the surface of a 3D metal [54]. The semiconductor 2DEG we employ at cryogenic temperatures exhibits larger kinetic inductance than 3D metals, leading to extraordinarily strong negative refraction at GHz frequencies, with an index as large as -700. This pronounced negative refractive index and corresponding reduction of the effective wavelength opens up a path to miniaturization in the science and technology of negative refraction.

The vision of creating negative refraction by exploiting the collective electron acceleration (inertia) effect, or kinetic inductance, was theoretically proposed with specific arrangements of 3D metallic nanoparticles [53]. Experimentally the inertia-based negative refraction was implied in the work where a particular guiding of surface plasmonic waves on the surface of a 3D metal led to negative refraction [54]; this cannot be explained without electrons' acceleration, since a defining component of plasmonic waves is the time-varying kinetic energy resulting from the acceleration of their constituent electrons.

The semiconductor 2DEG possesses much larger kinetic inductance than 3D bulk metals. Here we

[2]Large portions of this section are derived from a paper in publication by the author [14].

create negative index metamaterials by fully exploiting this large kinetic inductance of the semiconductor 2DEG, whose impact manifests in the extraordinarily large negative index, measured down to $n = -700$. This is two-orders-of-magnitude larger than n of $-5 \sim -1$ in the surface-plasmon-based negative refraction [54], indicating a far stronger role inertia plays in our 2DEG case. It is also far larger than the theoretical expectation based on the kinetic inductance of 3D metallic nanoparticles [53], which is orders-of-magnitude smaller than our 2D kinetic inductance (see Sec. 3.2.1.3 for a numerical comparison).

3.2.1 Device Structure and Description of Behavior

3.2.1.1 Device Structure

We choose a GaAs/AlGaAs 2DEG as a demonstrational platform, where electrons can accelerate for ~ 0.2 ns at 4 K without scattering, rendering their large kinetic inductance effect not masked by the scattering at and above GHz frequencies. Specifically, our metamaterial is a periodic array of mesa-etched 2DEG strips (Figs. 3.7a,b), each of which is tied to ground lines (labeled 'G') at its both ends via ohmic contacts. Each strip's width and length are denoted as W and l; the center-to-center distance between neighboring strips, or periodicity, is denoted as a. This metamaterial is excited by electromagnetic waves guided by the left signal line (labeled 'S'), which, flanked by the ground lines, forms an on-chip coplanar waveguide (CPW) with a 50-Ω characteristic impedance. This left signal line is extended up to over a few 2DEG strips on the left side of the metamaterial, with dielectric in between. The metamaterial's response is picked up by the right signal line (also labeled 'S') of another CPW on the right side of the metamaterial.

3.2.1.2 Principle of Operation

The excitation electromagnetic wave's electric fields, oscillating between the signal and ground lines of the left CPW, collectively accelerate electrons in the leftmost few 2DEG strips, producing currents along the strips. The resulting alterations of charge distribution in these strips will capacitively couple to neighboring strips to the right, accelerating electrons there. This process repeats to deliver an *effective wave* from left to right, *perpendicular* to the direction of the strips. From the circuit point of view, each 2DEG strip, along which electrons collectively accelerate with the resulting current lagging the

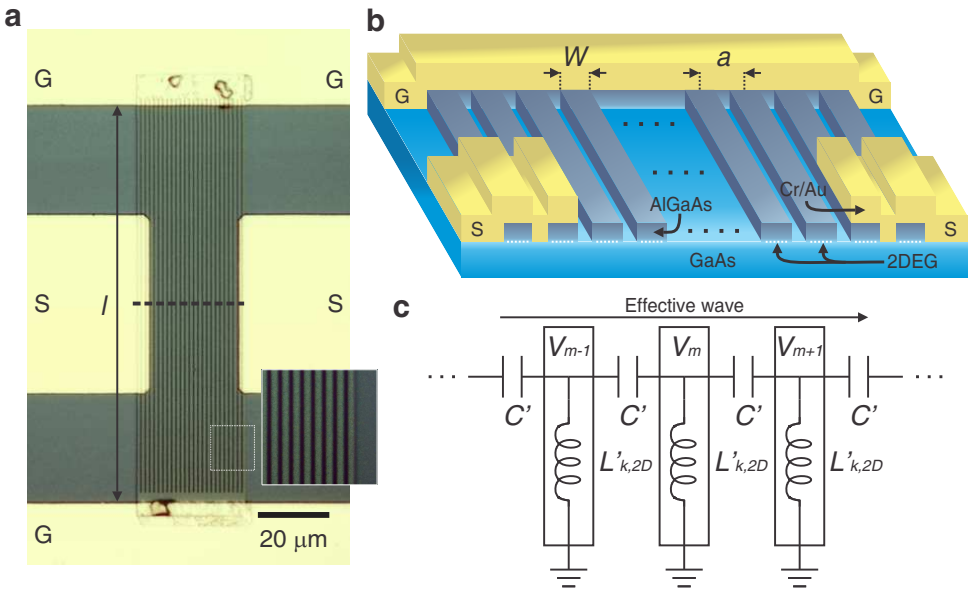

Figure 3.7: Description of device structure [14]. **a**, Optical image of a 2DEG strip array metamaterial prototype. Ground-Signal-Ground (GSG) on-chip CPWs direct electromagnetic waves to and from the metamaterial. The inset shows a zoomed-in portion of the strip array. In this specific prototype, $W = 1\ \mu$m, $l = 112\ \mu$m, and $a = 1.25\ \mu$m. **b**, Schematic perspective of the metamaterial (not drawn to scale), with the front face corresponding to the vertical cut through the dashed line in part **a**. **c**, Circuit description of the half structure of the metamaterial below or above the horizontal symmetry line along the wave propagation direction (see Sec. 3.2.2.1).

accelerating voltage by $90\ ^\circ$ according to Newton's law, acts as non-magnetic inductance of kinetic origin [20, 52]. This 2D kinetic inductance, $L'_{\mathrm{k,2D}}$, results from Newton's law of motion[3]:

$$L'_{\mathrm{k,2D}} = \frac{m^*}{n_0 e^2}\frac{l_{\mathrm{e}}}{W}, \tag{3.1}$$

where m^*, e, and n_0 are electrons' effective mass, charge, and density per unit area. l_{e}, which will be identified shortly, is the effective length of each strip within which electrons accelerate in response to the excitation. Our metamaterial is then an array of capacitively-coupled kinetic inductors (Fig. 3.7c; see Sec. 3.2.2.1), and may be likened to the left-handed transmission line [55–57], the array of capacitively-coupled magnetic inductors, which is known to be negatively refracting. From the physics viewpoint, our device uses extremely large 2DEG kinetic inductance, while the left-handed line relies on much smaller magnetic inductance.

To examine the negatively-refracting behavior of our device, we represent the effective wave with

[3]Note that we here use a lumped inductance as opposed to the per-unit-length inductance (Eq. (2.3)) that we have used so far; prime (') was added to the notation to signify this difference.

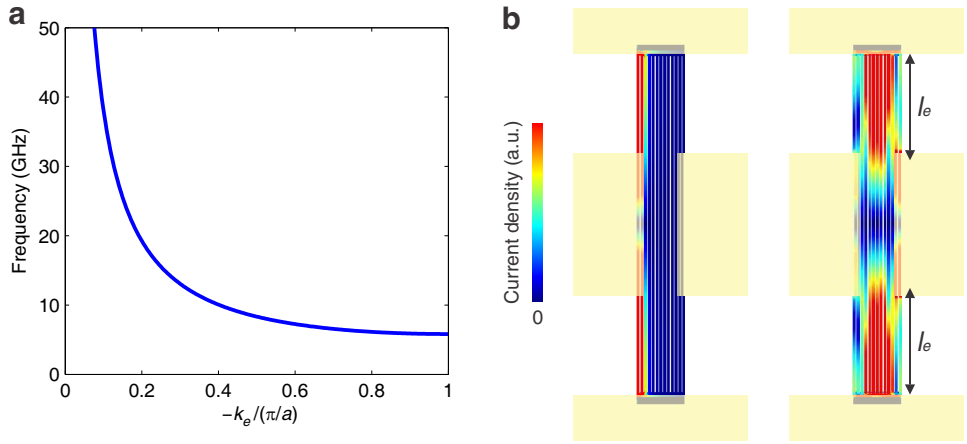

Figure 3.8: Theoretical dispersion relation and simulation [14]. **a**, Plot of $\omega(k) = \left[2\sqrt{L'_{\mathrm{k,2D}}C'}|\sin(k_\mathrm{e}a/2)|\right]^{-1}$, with $L'_{\mathrm{k,2D}} = 39$ nH and $C' = 4.6$ fF estimated for the structure measured for Fig. 3.12 ($W = 1\ \mu$m, $a = 1.25\ \mu$m, $l = 112\ \mu$m). Group and phase velocities, $d\omega/dk_\mathrm{e}$ and ω/k_e, have opposite signs, showing negative refraction; this occurs for both $k_\mathrm{e} > 0$ and $k_\mathrm{e} < 0$, but we only show the latter, in relevance to our measurements. **b**, Simulated current distributions below (left, 5 GHz) and above (right, 30 GHz) the cutoff frequency. Red and blue colors indicate high and low current density. Above the cutoff frequency, regions of high, constant current density are observed, from which the effective strip length l_e is estimated.

the voltage at the tip of the m-th kinetic inductor (Fig. 3.7c) as $V_m(t) \sim e^{i(\omega t - mk_\mathrm{e}a)}$, where ω is the angular frequency and k_e is the effective wavenumber. The standard circuit analysis of Fig. 3.7c yields a dispersion relation,

$$\omega(k) = \frac{\omega_\mathrm{c}}{|\sin(k_\mathrm{e}a/2)|}, \tag{3.2}$$

where $\omega_\mathrm{c} \equiv 1/2\sqrt{L'_{\mathrm{k,2D}}C'}$ is the cutoff frequency at the first Brillouin-zone boundary, $k_\mathrm{e} = \pm\pi/a$, and C' is the capacitance between adjacent strips over the effective length (see Sec. 3.2.2.1). For $\omega > \omega_\mathrm{c}$, the dispersion relation (Fig. 3.8a) clearly predicts negative refraction, as the $d\omega/dk_\mathrm{e}$ tangential slope (group velocity) and the ω/k_e slope (phase velocity) have opposite signs [51]. The cutoff behavior results from the metamaterial's high-pass nature, and can be also seen from the current distributions across the metamaterial below and above the cutoff frequency (Fig. 3.8b, right), simulated using an electromagnetic field solver (see Sec. 3.2.4). Beyond the cutoff frequency, the current is concentrated at the bottom and top regions of the strips, from which l_e can be estimated. Interestingly, while a *sheet* of 2DEG exhibits ordinary dispersion [10], its slicing into strips and acceleration of electrons along these strips *perpendicular* to the effective wave propagation causes negative refraction.

The dispersion relation of our metamaterial assumes the same form as that of the left-handed trans-

Table 3.1: Kinetic and magnetic inductances per unit length for 3D gold nanoparticles with various cross-sectional areas [14]. L_{m} is estimated assuming no other conductors nearby.

A	$L_{\mathrm{k,3D}}$	L_{m}	$L_{\mathrm{k,3D}}/L_{\mathrm{m}}$
1×1 nm^2	600 pH/μm	~ 1.4 pH/μm	430
5×5 nm^2	24 pH/μm	~ 1.2 pH/μm	20
10×10 nm^2	6.0 pH/μm	~ 1.0 pH/μm	6
20×20 nm^2	1.5 pH/μm	~ 0.9 pH/μm	1.7

mission line [55–57], but with the magnetic inductance replaced with the much larger 2DEG kinetic inductance; the 2DEG kinetic inductance is 1.25 nH/μm for a 1-μm wide 2DEG strip, which is $\sim$2,800 times larger than the same strip's magnetic inductance, 0.44 pH/μm (see Sec. 3.2.1.3 for details). The effective refractive index

$$n = -\frac{2c}{a\omega}\sin^{-1}\left(\frac{\omega_{\mathrm{c}}}{\omega}\right) \tag{3.3}$$

(c: free-space speed of light) derived from the dispersion has the maximum attainable magnitude of $\pi c/a\omega_{\mathrm{c}} = (2\pi c/a)\sqrt{L'_{\mathrm{k,2D}}C'}$, which is exceedingly large with the large 2DEG kinetic inductance, corresponding to the significant slowing of the effective wave.

3.2.1.3 Numerical Comparison of 2D and 3D Kinetic Inductance

The kinetic inductance is significantly more pronounced in semiconductor 2DEG than in 3D bulk metal. This is seen from numerical calculations in Tables 3.1 and 3.2, where the kinetic and magnetic inductances, L_{k} and L_{m}, are calculated for 3D gold nanoparticles [53] of varying cross-sectional area and for 2DEG strips of varying cross-sectional width. As can be seen, the 2DEG strip has both much larger absolute kinetic inductance and much larger kinetic-to-magnetic inductance ratio. In both quantities, typical cases compare with a factor of at least 100 or more, even when the 2DEC strip has much larger cross-sectional linear scale than the 3D gold nanoparticle. For the same cross-sectional linear scale ($\sqrt{A} = W$), the difference is even larger (*e.g.*, one can compare the bottom row of Table 3.1 to the top row of Table 3.2 where $\sqrt{A} = W = 20$ nm; L_{k} is a factor of 42,000 larger, and $L_{\mathrm{k}}/L_{\mathrm{m}}$ is a factor of 16,000 larger). Note that the kinetic-to-magnetic inductance ratio is a key factor determining the magnitude of the negative index, thus, these tables show the impact of creating a large negative index by utilizing the 2DEG kinetic inductance.

Table 3.2: Kinetic and magnetic inductances per unit length for GaAs/AlGaAs 2DEG strips with various cross-sectional widths [14]. $L_{\rm k}$ is calculated using the material parameters of the 2DEG used in this work. $L_{\rm m}$ is simulated with the Sonnet electromagnetic field solver, with no other conductors nearby, but inside a $400\ \mu{\rm m} \times 400\ \mu{\rm m} \times 1000\ \mu{\rm m}$ conducting box. Each strip is $400\ \mu{\rm m}$ long and is connected to the walls at the center of the $400\ \mu{\rm m} \times 1000\ \mu{\rm m}$ side walls. The magnetic inductance so simulated is frequency dependent, and the table presents the value averaged over $5 \sim 50$ GHz frequency range. This simulation is conservative, as in our actual device, $L_{\rm m}$ is even smaller due to neighboring strips (*e.g.*, simulated $L_{\rm m}$ for a 1-μm wide strip in the presence of neighboring strips is $0.44\ {\rm pH}/\mu{\rm m}$, as stated in Sec. 3.2.1.2, instead of $1.4\ {\rm pH}/\mu{\rm m}$).

W	$L_{\rm k,2D}$	$L_{\rm m}$	$L_{\rm k,2D}/L_{\rm m}$
20 nm	$62.5\ {\rm nH}/\mu{\rm m}$	$\sim 2.3\ {\rm pH}/\mu{\rm m}$	27,000
100 nm	$12.5\ {\rm nH}/\mu{\rm m}$	$\sim 2.0\ {\rm pH}/\mu{\rm m}$	6,300
500 nm	$2.50\ {\rm nH}/\mu{\rm m}$	$\sim 1.6\ {\rm pH}/\mu{\rm m}$	1,600
1000 nm	$1.25\ {\rm nH}/\mu{\rm m}$	$\sim 1.4\ {\rm pH}/\mu{\rm m}$	890

3.2.2 Dispersion Relation and Effective Refractive Index

3.2.2.1 Derivation of Dispersion Relation

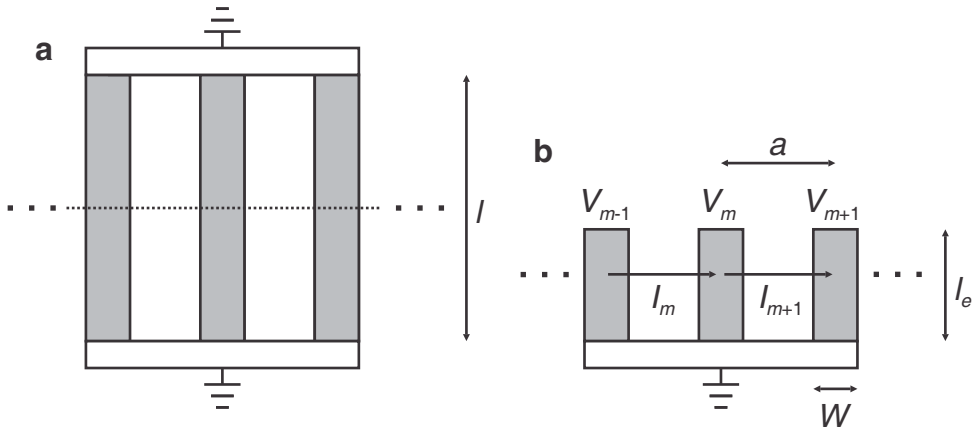

Figure 3.9: Schematic model of 2DEG strip array [14]. **a**, 2DEG strip array. Dark regions indicate 2DEG. **b**, Half-structure cut at the horizontal symmetry line. Only the effective length $l_{\rm e}$ of each 2DEG strip is relevant.

This section derives the dispersion relation for the *effective wave* traveling through the 2DEG strip array (Fig. 3.9a). Since no current flows *vertically* through the center of any strip due to symmetry, only the lower/upper half below/above the horizontal symmetry line can be considered (Fig. 3.9b). In this half-structure, each strip has width W, effective length[4] $l_{\rm e}$, and periodicity a. Each strip in this

[4]Because the field distribution of the guided modes in the CPWs is concentrated between the signal and ground lines, not the total geometric strip length l but the effective length $l_{\rm e}$ within which the guided modes accelerate electrons is relevant. To a good approximation, the distance between the signal and ground lines of the CPWs can be used for $l_{\rm e}$, as shown in Fig. 3.8b, right.

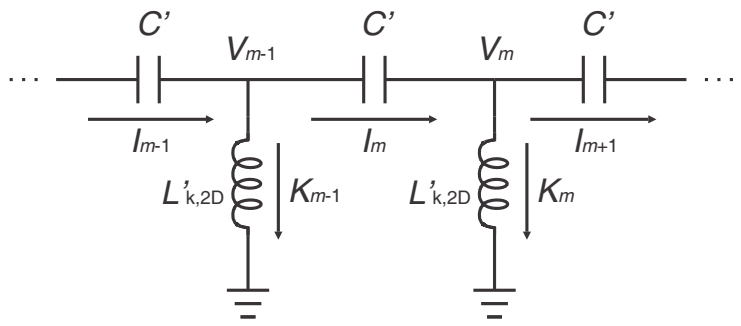

Figure 3.10: Circuit model for the 2DEG strip array [14].

half-structure may be considered as a lumped kinetic inductance $L'_{\mathrm{k,2D}}$ of Eq. (3.1). We ignore the loss for simplicity. The interactions between two neighboring strips within the effective length can be modeled using a lumped capacitance C' between them (we only consider the nearest-neighbor couplings for simplicity), given by [58]

$$C' = \kappa\epsilon_0 l_{\mathrm{e}} \frac{K(\sqrt{1-u^2})}{K(u)}, \tag{3.4}$$

where $u \equiv (a-W)/(a+W)$, κ is the effective dielectric constant of the surroundings, and $K(u)$ is the complete elliptic integral of the first kind:

$$K(u) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-u^2\sin^2\theta}}. \tag{3.5}$$

C' does not vary simply with the inter-strip distance, $a-W$, but varies in a more complicated fashion with u, due to the confinement into two dimensions. The strip array then can be thought of as a capacitively-coupled inductor array (Fig. 3.10). This lumped circuit model is a good approximation, for the plasmonic wavelength along the strip is mostly much longer than l_{e} (see the following Sec. 3.2.2.2 for a more detailed discussion of this approximation).

The effective wave is represented by voltage $\{\cdots, V_{m-1}, V_m, V_{m+1}, \cdots\}$ at the tips of the kinetic inductors (Figs. 3.9 and 3.10). For the m-th capacitor's current, we have

$$C' \frac{d}{dt}(V_{m-1} - V_m) = I_m. \tag{3.6}$$

According to the Kirchhoff's current law, we have

$$I_m - I_{m+1} = K_m, \tag{3.7}$$

where K_m is the current flowing through the m-th kinetic inductor, and is related to V_m via

$$V_m = L'_{\mathrm{k,2D}} \frac{dK_m}{dt}. \tag{3.8}$$

By combining Eqs. (3.6) to (3.8), we obtain the wave equation,

$$\frac{1}{L'_{\mathrm{k,2D}} C'} V_m = \frac{d^2}{dt^2} \left(V_{m-1} + V_{m+1} - 2V_m \right). \tag{3.9}$$

Plugging $V_m = V_0 e^{i(\omega t - k_\mathrm{e} m a)}$ into Eq. 3.9, where k_e is the effective wavenumber and ω is the angular frequency, we obtain the dispersion relation,

$$\omega(k) = \frac{1}{2\sqrt{L'_{\mathrm{k,2D}} C'}} \frac{1}{|\sin(k_\mathrm{e} a/2)|} = \frac{\omega_\mathrm{c}}{|\sin(k_\mathrm{e} a/2)|}, \tag{3.10}$$

where

$$\omega_\mathrm{c} \equiv \frac{1}{2\sqrt{L'_{\mathrm{k,2D}} C'}} \tag{3.11}$$

is the cutoff frequency corresponding to $k_\mathrm{e} = \pm\pi/a$, the first Brillouin-zone boundary. Eq. (3.10) exhibits opposite signs of $d\omega/dk_\mathrm{e}$ and ω/k_e (group and phase velocities) for $\omega > \omega_\mathrm{c}$, showing negative refraction. This holds for both $k_\mathrm{e} > 0$ and $k_\mathrm{e} < 0$, but in Fig. 3.8a, we only show $k_\mathrm{e} < 0$, as $k_\mathrm{e} > 0$ is irrelevant to the measurements of this work (Secs. 3.2.3.2 and 3.2.3.3). The refractive index is in general[5]

$$n = \frac{\mathrm{sgn}(d\omega/dk)}{\mathrm{sgn}(\omega/k)} \frac{|k|c}{\omega}, \tag{3.12}$$

[5]In our measurements, if the device under test is negatively [positively] refracting, $k_\mathrm{e} < 0$ [$k_\mathrm{e} > 0$], so the index expression reduces to $n = k_\mathrm{e} c/\omega$.

where $\mathrm{sgn}(x) \equiv x/|x|$. By using Eq. (3.10) here, we obtain

$$n = -\frac{2c}{a\omega}\sin^{-1}\Big(\frac{\omega_{\mathrm{c}}}{\omega}\Big). \tag{3.13}$$

3.2.2.2 Plasmonic Wave Consideration

So far we have represented the effect of the collective electron acceleration along the 2DEG strip as a lumped kinetic inductor. More generally, this collective electron acceleration is part of the collective plasmonic wave excitation *along* the 2DEG strip (not to be confused with the effective wave traversing *perpendicularly* to the strips). The plasmonic wave behaviors within an individual 2DEG strip can be understood from the circuit point of view, by treating the strip as a plasmonic transmission line (Sec. 2.1). The distributed kinetic inductances account for the time-varying kinetic energy (acceleration) of electrons; the kinetic inductance per unit length is given by $L_{\mathrm{k,2D}} \equiv L'_{\mathrm{k,2D}}/l_{\mathrm{e}}$. The distributed capacitances with the per-unit-length value of C account for the time-varying potential energy due to the Coulomb interaction and degeneracy pressure within the 2DEG strip; in our frequency range, the degeneracy pressure effect is negligible, and C is of purely electrostatic nature. These distributed capacitances C in the plasmonic line are connected to ground at their one end, and should not be confused with the inter-strip capacitance C' of Eq. (3.4) that plays a role in propagating the effective wave through the strip array. C is generally frequency dependent, and depends on the presence/absence and detailed configuration of grounded conducting objects nearby. The plasmonic velocity is $v_{\mathrm{p}} = 1/\sqrt{L_{\mathrm{k,2D}}C}$. The plasmonic line's characteristic impedance is $Z_0 = \sqrt{L_{\mathrm{k,2D}}/C}$.

As discussed shortly, in our work, the plasmonic wavelength λ_{p} along the strip is mostly much longer than the strip's effective length l_{e}, or equivalently $k_{\mathrm{p}}l_{\mathrm{e}} \ll 1$ (k_{p}: plasmonic wavenumber along the direction of the strip). In this long wavelength regime, our 2DEG plasmonic line, whose one end is grounded via finite ohmic contact resistance R_{c} that is much smaller than Z_0, acts as a lumped kinetic inductor. This can be seen by calculating the input impedance Z_{in} of the 2DEG strip at its non-grounded end. To begin with, with $R_{\mathrm{c}} \ll Z_0$, we have [25]

$$Z_{\mathrm{in}} = Z_0\frac{R_c + iZ_0\tan k_{\mathrm{p}}l_{\mathrm{e}}}{Z_0 + iR_c\tan k_{\mathrm{p}}l_{\mathrm{e}}} \approx iZ_0\tan k_{\mathrm{p}}l_{\mathrm{e}}. \tag{3.14}$$

Now with $k_{\mathrm{p}} l_{\mathrm{e}} \ll 1$, $\tan k_{\mathrm{p}} l_{\mathrm{e}} \approx k_{\mathrm{p}} l_{\mathrm{e}}$, thus, Z_{in} further reduces to

$$Z_{\mathrm{in}} \approx i Z_0 k_{\mathrm{p}} l_{\mathrm{e}} = i Z_0 \frac{\omega}{v_{\mathrm{p}}} l_{\mathrm{e}} = i\omega L_{\mathrm{k,2D}} l_{\mathrm{e}}. \tag{3.15}$$

This shows that the 2DEG strip behaves as a lumped kinetic inductor $L_{\mathrm{k,2D}} l_{\mathrm{e}}$, which is no more than $L'_{\mathrm{k,2D}}$ of Eq. (3.1). In fact, even when a shorter wavelength is considered without satisfying $k_{\mathrm{p}} l_{\mathrm{e}} \ll 1$, as far as $k_{\mathrm{p}} l_{\mathrm{e}} < \pi/2$, Z_{in} is still inductive with the inductance value even larger than $L_{\mathrm{k,2D}} l_{\mathrm{e}}$ as seen from Eq. (3.14), thus, the strip array will maintain its negatively refracting behavior. By contrast, if $k_{\mathrm{p}} l_{\mathrm{e}} > \pi/2$, Z_{in} would become capacitive, making the negative refraction vanish in the strip array device. In between these two regimes, *i.e.*, at $k_{\mathrm{p}} l_{\mathrm{e}} = \pi/2$ ($l_{\mathrm{e}} = \lambda_{\mathrm{p}}/4$), plasmonic standing wave resonance occurs.

To compare λ_{p} to l_{e} in our frequency range, we estimate λ_{p} by resorting to the ungated 2D plasmonic dispersion relation (Eq. (2.40)), which we rewrite here in terms of m^* and n_0 as

$$\omega = \sqrt{\frac{n_0 e^2 k_{\mathrm{p}}}{2\kappa\epsilon_0 m^*}}, \tag{3.16}$$

where $n_0 = 1.9 \times 10^{11}/\mathrm{cm}^2$, $m^* = 0.067 m_{\mathrm{e}}$, and κ is the effective dielectric constant given by $\kappa_{=}(\kappa_{\mathrm{GaAs}} + \kappa_{\mathrm{Air}})/2 = 6.95$, in which we safely ignore the very thin AlGaAs layer (71 nm, including a thin GaAs cap) [34]. Let us consider the longest strip ($l = 112\ \mu\mathrm{m}$; $l_{\mathrm{e}} = 31\ \mu\mathrm{m}$) employed in this work. At 50 GHz, $\lambda_{\mathrm{p}} \sim 400\ \mu\mathrm{m} \sim 3.6l \sim 13 l_{\mathrm{e}}$; at 25 GHz, $\lambda_{\mathrm{p}} \sim 1600\ \mu\mathrm{m} \sim 14l \sim 50 l_{\mathrm{e}}$. So even at 50 GHz and even with the longest strip, $k_{\mathrm{p}} l_{\mathrm{e}} \sim 0.5$ and $\tan k_{\mathrm{p}} l_{\mathrm{e}} \sim 0.54 \sim k_{\mathrm{p}} l_{\mathrm{e}}$; thus, even in this conservative case, Eq. (3.15) is valid, justifying the treatment of the 2DEG as a lumped kinetic inductance given by Eq. (3.1). The first plasmonic standing wave resonance condition $l_{\mathrm{e}} = \lambda_{\mathrm{p}}/4$ is beyond our frequency range. To be even more conservative (albeit not realistic), even with the actual geometric length l, the first plasmonic standing wave condition $l = \lambda_{\mathrm{p}}/2$ is still beyond our frequency range.

3.2.2.3 Dependency of n on $L'_{\mathrm{k,2D}}$

With Eqs. (3.10) and (3.13), we examine how the very large $L'_{\mathrm{k,2D}}$ yields the extraordinarily strong negative refraction, in comparison to the conventional left-handed transmission line [55–57], which has essentially the same dispersion expression, but with magnetic inductance L_{m} in the place of $L_{\mathrm{k,2D}}$.

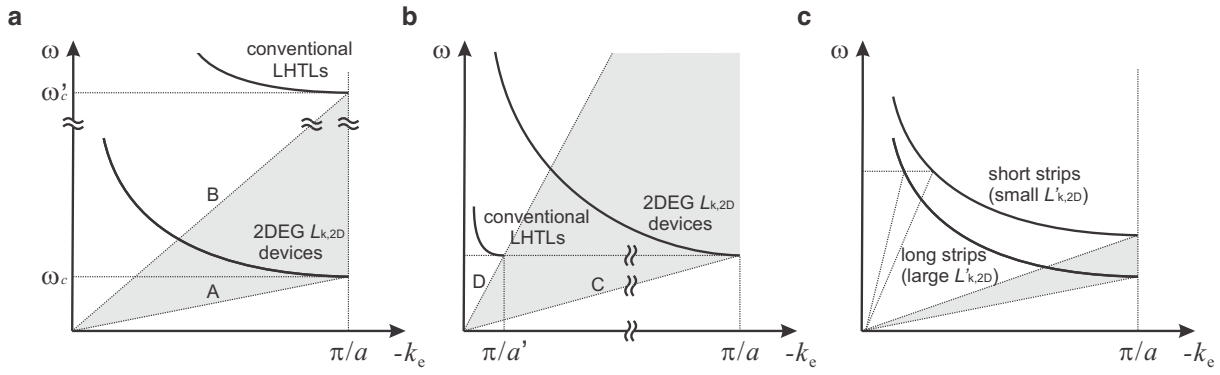

Figure 3.11: Comparison of LHTL and 2DEG strip array [14]. **a**, Comparison of the dispersion relations for conventional left-handed transmission lines (LHTLs) and our 2DEG strip array metamaterials for the same a. **b**, Comparison of the dispersion relations for conventional LHTLs and our 2DEG strip array metamaterials for the same $\omega_{\rm c}$. **c**, Comparison of our metamaterials with shorter and longer strips for the same a. Plots are not drawn to scale.

Imagine a left-handed transmission line with similar physical dimensions to our 2DEG strip array. For example, one can think of the same strip array as our metamaterial with the same geometric parameters (a, W, and l), but with $L_{\rm m}$ replacing $L_{\rm k,2D}$. Since this left-handed line and our metamaterial have the same l, $L_{\rm k,2D} \gg L_{\rm m}$. For example, $L_{\rm k,2D}$ is 1.25 nH/μm for a 1-μm wide 2DEG strip, which is $\sim$2,800 times larger than the same strip's $L_{\rm m} \sim 0.44$ pH/μm. In this case, the left-handed line's cutoff frequency $\omega_{\rm c}'$ is far larger than our metamaterial's cutoff frequency $\omega_{\rm c}$ (Eq. (3.11)), while a remains the same. Their dispersion relations are illustrated in Fig. 3.11a. The slope of a straight line connecting the origin and a $(\omega, k_{\rm e})$ point is the phase velocity $\omega/k_{\rm e}$ at frequency ω. Entering the shaded region forbidden for the left-handed line, our metamaterial achieves a range of $\omega/k_{\rm e}$ values far smaller than the minimum $\omega/k_{\rm e}$ of the left-handed line. Equivalently, in this region, our metamaterial exhibits a range of negative refractive index values much larger in magnitude than the maximum negative index of the left-handed line. The largest attainable negative index for our metamaterial (which corresponds to line **A**, Fig. 3.11a), $n = -\pi c/a\omega_{\rm c}$, is much larger in magnitude than that for the left-handed line (line **B**), $n = -\pi c/a\omega_{\rm c}'$, as $\omega_{\rm c} \ll \omega_{\rm c}'$.

Alternatively, one can consider a left-handed line and our metamaterial with $L_{\rm m} = L_{\rm k,2D}$; since the magnetic inductance per unit length is much smaller than the kinetic inductance per unit length, $L_{\rm m} = L_{\rm k,2D}$ will mandate much larger physical dimensions for the magnetic inductors. In this scenario, $\omega_{\rm c}$ can be made the same between our metamaterial and the left-handed line, but due to the much larger physical

dimensions of the magnetic inductors, the left-handed line's periodicity a' will be much larger than our metamaterial's periodicity a. The dispersion relations for these two devices are illustrated in Fig. 3.11b. In the shaded area forbidden for the left-handed line, our metamaterial exhibits a range of negative refractive index values much larger in magnitude than the maximum negative index of the left-handed line. The largest attainable negative index for our metamaterial (line **C** in Fig. 3.11b) $n = -\pi c/a\omega_{\rm c}$ is much larger in magnitude than that for the left-handed line (line **D** in Fig. 3.11b) $n = -\pi c/a'\omega_{\rm c}$, as $a' \gg a$.

By the same token, the effect of the change in l (or $l_{\rm e}$) between our 2DEG metamaterials can be examined graphically as well. With the a and W fixed, longer $l_{\rm e}$ causes C' and $L'_{\rm k,2D}$ to increase, resulting in lower $\omega_{\rm c}$. Fig. 3.11c shows the dispersion curves for our metamaterial with the same a but differing $\omega_{\rm c}$. Only the longer-strip metamaterial can access the shaded region, in which it attains negative indices larger in magnitude than the maximum negative index of the shorter-strip metamaterial; the longer-strip metamaterial has a larger maximum attainable negative refractive index. On the other hand, at the same frequency accessible for both metamaterials, the shorter-strip metamaterial exhibits a larger negative refractive index.

3.2.3 Measurement and Analysis

Microwave scattering experiments with on-chip probing confirm the extraordinarily strong negative refraction. The reflection of an electromagnetic wave launched onto the left on-chip CPW and its transmission onto the right on-chip CPW after propagation through the metamaterial are measured over $1 \sim 50$ GHz using a vector network analyzer. Propagation delays in the two on-chip CPWs and parasitic couplings between them bypassing the metamaterial were separately measured and de-embedded; from the resulting transmission and reflection coefficients, s_{21} and s_{11}, at each measurement frequency, we extract, using a well-established prescription [59–63], the effective wave's phasor change $e^{-ik_{\rm e}d}$ due purely to propagation across the metamaterial of distance d (see Sec. 3.2.3.2). Fig. 3.12a shows the f-$k_{\rm e}$ dispersion so obtained at temperatures 4.2, 10, and 20 K for a 13-strip metamaterial with $W = 1$ μm, $l = 112$ μm, $l_{\rm e} = 31$ μm, and $a = 1.25$ μm. Since the measured s_{21} and s_{11} parameters set the left-to-right energy propagation direction (*i.e.*, group velocity direction) as the positive reference direction, if our metamaterial is negatively refracting, the extracted $k_{\rm e}$'s sign will be minus with no ambiguity,

which is indeed seen in Fig. 3.12a. Negative refraction is also consistently confirmed in Fig. 3.12a by the fact that the $d\omega/dk_{\mathrm{e}}$ tangential slope and the ω/k_{e} slope have opposite signs above the 12-GHz cutoff frequency. This measured dispersion, including the cutoff frequency, is different in details from the calculation that uses lumped circuit elements, ignores losses due to electron scattering in the 2DEG strips and ohmic contacts, and considers only nearest capacitive couplings (Fig. 3.8a). But the same essence prevails, demonstrating negative refraction. The dark area in Fig. 3.12a, where the distinctively spurious behavior of the dispersion appears, is indicative of the cutoff region, which is irrelevant to the device operation (see Sec. 3.2.3.2). From this f-k_{e} dispersion, we obtain the effective refractive index via $n = k_{\mathrm{e}}c/\omega$, whose real part is as large as -500 (Fig. 3.12b). This large negative index, which is difficult to achieve, if not impossible, with *magnetic* inductance [44–46, 55–57, 64, 65], enables dramatic device miniaturization and can facilitate ultra-subwavelength localization. The same measurements performed on the 2DEG strip array, but with energy propagation *along* the strips, yield positive refraction, further highlighting our negative refraction strategy (see Sec. 3.2.3.3).

$|\mathrm{Re}(n)|$ decreases with frequency (Fig. 3.12b), because as frequency becomes higher, adjacent strips are coupled more capacitively, smearing the electron acceleration effect within separated strips. Fig. 3.12c shows the figure of merit, $|\mathrm{Re}(n)/\mathrm{Im}(n)|$, which here reflects losses due to electron scatterings in the 2DEG strips and ohmic contacts. It is around 2 in good part of the negative refraction region, similarly to negative refraction devices using metals at optical frequencies [64, 65]. Figs. 3.12a-c show that negative refraction behaviors are essentially consistent regardless of temperature (4.2, 10, 20 K), indicating that the degree of electron scatterings in the 2DEG strips and ohmic contacts remains largely the same within this temperature range, not masking the inertia effect. In Fig. 3.12c, the figure of merit is largest at 10 K instead of 4.2 K, but these variations with respect to temperature arise mostly by inconsistent probe landings during multiple calibration steps, which are executed for measurements at each temperature. Fluctuations at high frequencies, *e.g.*, in Fig. 3.12c, are also due to imperfect calibration.

At 297 K, electron scattering in each 2DEG strip becomes severe to completely mask the acceleration effect. Equivalently, the strip's ohmic resistance becomes far larger (*ca.* $100\ \mathrm{k}\Omega$) than the impedance of its kinetic inductance. The strip array then becomes essentially open-circuited, causing the signal to be mostly reflected. This reflection can be seen from $|s_{11}| \sim 1$ measured at 297 K, which differs from $|s_{11}|$ at cryogenic temperatures where the strip array exhibits negative refraction (Fig. 3.12d). The

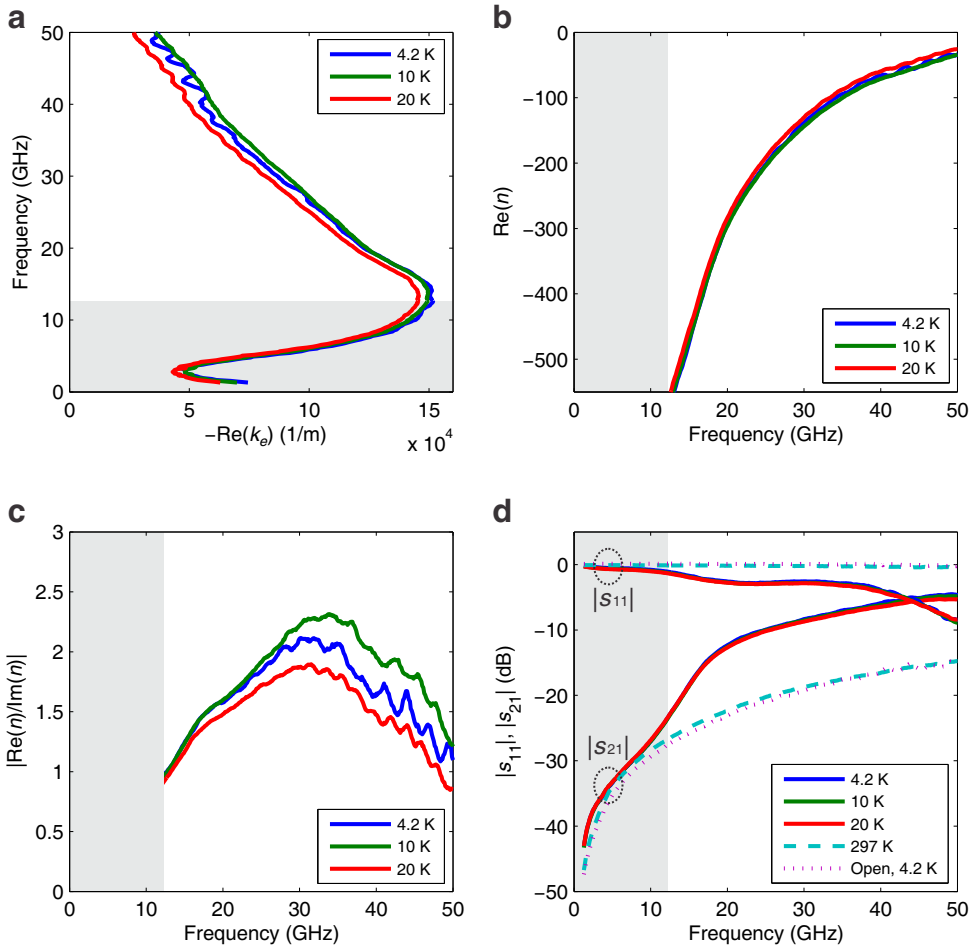

Figure 3.12: Temperature-dependent measurements [14]. **a**, Dispersion of the 13-strip metamaterial at 4.2, 10, and 20 K. The dark region is indicative of the cutoff behavior. **b**, $\mathrm{Re}(n)$ vs. frequency. **c**, Figure of merit $|\mathrm{Re}(n)/\mathrm{Im}(n)|$ vs. frequency. **d**, $|s_{11}|$ and $|s_{21}|$ of the metamaterial at 4.2, 10, 20, and 297 K. Also shown are $|s_{11}|$ and $|s_{21}|$ of the open device at 4.2 K. Unlike parts **a**, **b**, and **c**, these are raw s-parameters without de-embedding, to show the parasitic coupling between the two CPWs.

transmission $|s_{21}|$ becomes smaller at 297 K due also to the open-circuit behavior, but is not inappreciable (Fig. 3.12d). To understand this, we fabricate exactly the same structure as the previous device, but without the strip array, thus, creating an actual open circuit between the two on-chip CPWs. $|s_{21}|$ of this open device at 4.2 K closely resembles $|s_{21}|$ of the strip array device at 297 K (Fig. 3.12d). This demonstrates that the latter is due largely to the parasitic coupling between the two CPWs bypassing the strip array, confirming the open-circuit nature of the strip array at 297 K (in fact, the phases of s-parameters are also much similar between the 4.2-K open device and 297-K strip array; see Sec. 3.2.3.4). These results constitute another confirmation that our negative refraction at cryogenic temperatures is due to the kinetic inductance. Also in the most of the dark region (Fig. 3.12d), $|s_{21}|$ of the metamaterial even at cryogenic temperatures is much similar to that of the open device, confirming the cutoff nature in that region.

To further examine the impact of kinetic inductance on negative refraction, we measure a new set of devices with varying geometric parameters. Comparison of devices with different l (thus different l_{e}) for the same W and a is especially illuminating; it scales each of $L'_{\mathrm{k,2D}}$, C', and ω_{c} proportionally to l_{e}, affecting the index $n = -(2c/a\omega)\sin^{-1}(\omega_{\mathrm{c}}/\omega)$ with only one parameter, ω_{c}. Specifically, a longer-strip device, with larger $L'_{\mathrm{k,2D}}$ and C', and smaller ω_{c}, will have a larger maximum attainable negative index magnitude, $\pi c/a\omega_{\mathrm{c}}$, reaching the frequency region forbidden for a shorter-strip device, while the shorter-strip device assumes a larger negative index at the same frequency, as discussed in Sec. 3.2.2.3. This clear-cut property evidently emerges in measurements of a pair of $a = 1.25\ \mu\mathrm{m}$ devices with differing l ($112\ \mu\mathrm{m}$ vs. $52\ \mu\mathrm{m}$), or differing l_{e} ($31\ \mu\mathrm{m}$ vs. $14\ \mu\mathrm{m}$) (Fig. 3.13a). This property is affirmed again with two additional pairs of devices (Fig. 3.13b), where the index is as large as -700.

Altering the periodicity a, which in general may have to be accompanied by altering W, affects the index $n = -(2c/a\omega)\sin^{-1}(\omega_{\mathrm{c}}/\omega)$ in a more complicated manner, due to simultaneous changes in a and ω_{c}. For $l = 112\ \mu\mathrm{m}$, as we decrease [a, W] from [$1.5\ \mu\mathrm{m}$, $1\ \mu\mathrm{m}$] to [$1.25\ \mu\mathrm{m}$, $1\ \mu\mathrm{m}$] to [$0.75\ \mu\mathrm{m}$, $0.6\ \mu\mathrm{m}$] with the first reduction increasing C' by 1.2 times with $L'_{\mathrm{k,2D}}$ unchanged, and second reduction increasing $L'_{\mathrm{k,2D}}$ by 1.7 times with C' unchanged, ω_{c} does not vary as significantly as a, due to ω_{c}'s square-root dependence on $L'_{\mathrm{k,2D}}$ and C'. Thus, a smaller periodicity will yield a larger negative index for the same frequency away from the cutoff regions, as evident in measurements (Fig. 3.13c). In these a and W variations, the magnitude of the characteristic impedance, thus, impedance mismatch, is varied,

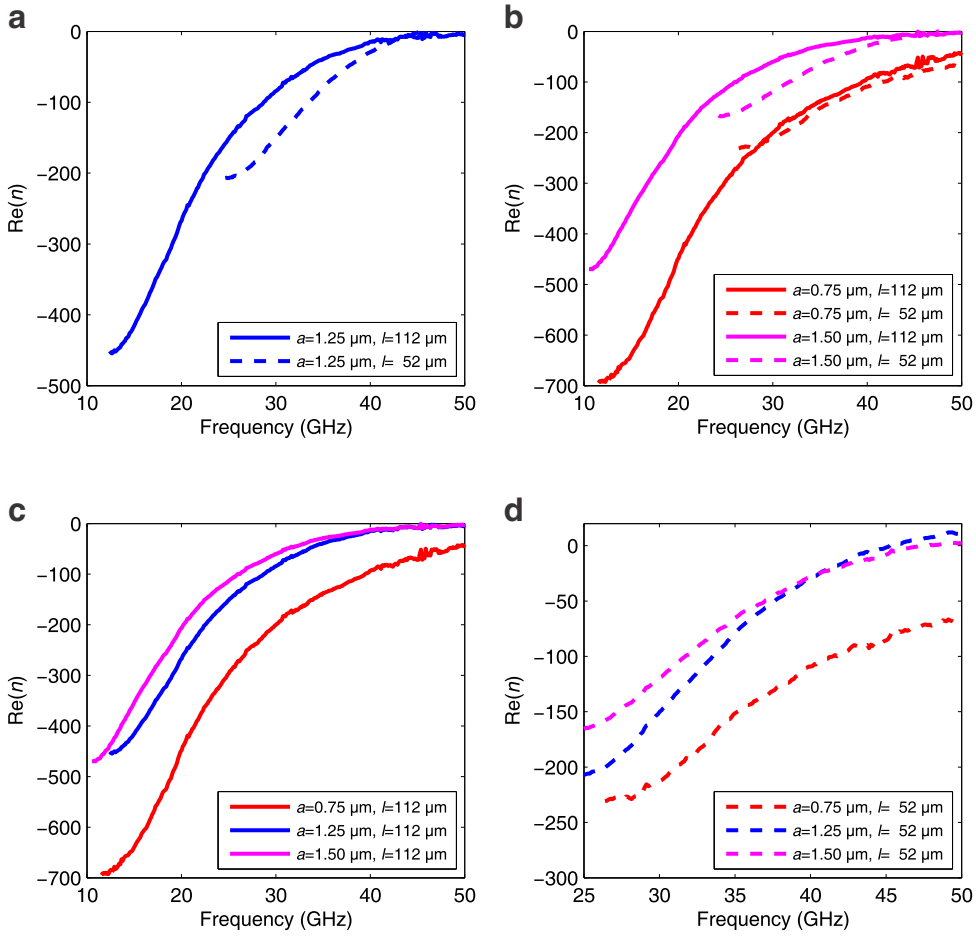

Figure 3.13: Geometry-dependent measurements [14]. **a**, $\mathrm{Re}(n)$ for a pair of 13-strip metamaterials with $a = 1.25\ \mu\mathrm{m}$ and $W = 1\ \mu\mathrm{m}$, but with differing l. **b**, $\mathrm{Re}(n)$ for another pair of 13-strip metamaterials with $a = 0.75\ \mu\mathrm{m}$ and $W = 0.6\ \mu\mathrm{m}$, for differing l; and for the third pair of 13-strip metamaterials with $a = 1.5\ \mu\mathrm{m}$ and $W = 1$, for differing l. **c**, $\mathrm{Re}(n)$ for $l = 112\ \mu\mathrm{m}$ with differing a and W. **d**, $\mathrm{Re}(n)$ for $l = 52\ \mu\mathrm{m}$ with differing a and W. Parts **c** and **d** are rearrangements of the data in **a** and **b** to facilitate the comparisons for the same l. Each device result is shown above its respective cutoff frequency.

yielding imperfect de-embedding of the parasitic couplings near cutoffs, obscuring the cutoff behaviors. The tendency of larger negative index for smaller periodicity is seen again for $l = 52$ μm with the same variations of a and W (Fig. 3.13d), except the data crossing between $a = 1.25$ μm and $a = 1.5$ μm devices, which we suspect arises from the impedance mismatch variation.

3.2.3.1 Frequency Scaling

The exceedingly strong inertia-based negative refraction demonstrated here requires a solid-state platform with very large kinetic inductance and low electron scattering, for which we employed the GaAs/AlGaAs 2DEG at cryogenic temperatures. Scaling the 2DEG metamaterial into higher frequencies by simultaneous reduction of the strip length and periodicity (Fig. 3.13) would relax the condition for electron scattering time and temperature; demonstration [66] of THz plasmonic devices at room temperature with GaAs/AlGaAs 2DEG bodes well for this direction. Graphene, another form of 2D conductor with high mobility at room temperature [67], may offer another potential platform for negative refraction at THz frequencies based on the similar kinetic approach; while individual electrons in graphene act as massless particles, they still possess kinetic energy and can exhibit plasmonic behaviors with implicit kinetic inductance; in fact, THz light-plasmon coupling has been recently observed at room temperature [12]. Graphene would have similar dimensions and frequency scaling properties as our GaAs 2DEG devices, because graphene's kinetic inductance with a feasible doping density is expected to be a few hundred pH/square [22], which is on the same order of magnitude as the kinetic inductance used in this work.

Measurements in Fig. 3.13 show, in keeping with the theory, that reducing the strip length l increases ω_c, but reduces $|n|$, and this $|n|$-reduction can be offset by decreasing the periodicity a. In fact, it is feasible to increase ω_c while maintaining a similar range of $|n|$, by reducing both l and a in a particular fashion, which we discuss now using the maximum attainable refractive index magnitude, $|n_{\max}| \equiv \pi c/(a\omega_c)$.

If l is reduced by a factor of[6] α (< 1), and a and W are reduced by a factor of β (< 1), ω_c is altered by a factor of $\sqrt{\beta}/\alpha$ (Eqs. (3.1), (3.4), and (3.11)) and $|n_{\max}|$ is changed by a factor of $\alpha/\beta^{3/2}$. To keep $|n_{\max}|$ constant, we must have $\alpha = \beta^{3/2}$; ω_c will then increase by a factor of $1/\beta = \alpha^{-2/3}$. As our 13-strip metamaterial with $l = 112$ μm, $W = 1$ μm, and $a = 1.25$ μm has a cutoff frequency of $\sim$10 GHz,

[6] To be rigorous, one must consider the scaling of l_e, but here we seek to provide the essence simply.

to scale up the cutoff frequency to 1 THz while keeping $|n_{\max}|$, $\beta = 1/100$ and $\alpha = 1/1000$. Thus this THz device would have $l = 112$ nm, $W = 10$ nm, and $a = 12.5$ nm. These W and a dimensions are at the verge of what is feasible with the current top-down fabrication technologies. In going to the 1-THz device, if we allow $|n_{\max}|$ to reduce by a factor of 5 (with which $|n_{\max}|$ is still $\sim$100), $\beta = 0.05$ and $\alpha = 0.00224$, yielding $l = 251$ nm, $W = 50$ nm, and $a = 62.5$ nm, which are well within the capability of current top-down fabrication technologies.

This high-frequency scaling is a theoretical projection; actual implementation of such small devices is expected to face various challenges. For example, the edge depletion effect, which is not a significant issue in the current work, will become pronounced when the 2DEG strip becomes narrower, as the edge depletion portion becomes an increasingly significant fraction of the strip, thus mandating careful design considerations. Coupling of the THz electromagnetic waves into such small negative refraction devices can pose another challenge, which one may be able to overcome by using wide-band log periodic antennae or other forms of electromagnetic structures [68]. At the same time, the small structures might offer some interesting avenues to explore. For example, as the 2DEG strip width approaches tens of nanometers, the 2DEG would start exhibiting 1D quantum wire behaviors (especially if it is operated at low temperatures, although we suggest the high-frequency scaling as a means to potentially achieve the room temperature operation), with 1D kinetic inductance quantized per each conduction channel. Exploiting this 1D kinetic inductance, if observable, for negative refraction would offer an interesting opportunity.

3.2.3.2 Extraction of n from s-Parameters

From the final set of s_{21} and s_{11} parameters at each measurement frequency after calibration and de-embedding of parasitic signals, we extract the effective wavenumber k_{e} inside the metamaterial by using the standard prescription detailed in Refs. [59] and [60]. Specifically, k_{e} is given by[7]

$$\begin{aligned} e^{ik_{\mathrm{e}}d} &= X \pm i\sqrt{1 - X^2}, \\ X &= \frac{1 - s_{11}^2 + s_{21}^2}{2s_{21}}. \end{aligned} \tag{3.17}$$

[7]These formulas directly from Refs. [59] and [60] use the phasor convention of $e^{i(k_{\mathrm{e}}x - \omega t)}$, which differs from the phasor convention $e^{i(\omega t - k_{\mathrm{e}}x)}$ our vector network analyzer uses. For the extraction to be consistent with our phasor convention, we actually use s_{11}^* and s_{21}^* in these formulas, which leads to exactly the same physical results [69].

Due to causality and analyticity, k_{e} can be uniquely determined despite the apparent multiplicity of signs and branches [59, 60]. This extraction connects the s_{21} and s_{11} parameters at any given measurement frequency f to the effective wavenumber k_{e} at the same frequency, thus, leading to the f-k_{e} dispersion curve, such as Fig. 3.12a. The measured s_{21} and s_{11} parameters set the left-to-right energy propagation direction (*i.e.*, group velocity direction) as the positive reference direction. Therefore, if the device under test is negatively [positively] refracting, the extracted k_{e}'s sign will be minus [plus], with no arbitrariness. For any physically meaningful situation, the refraction sign decision based on the sign of the extracted k_{e} will be consistent with the refraction sign decision made by examining the extracted dispersion curve as to whether the $d\omega/dk_{\mathrm{e}}$ tangential slope (group velocity) and ω/k_{e} slope (phase velocity) have the opposite or same signs. In sum, a negative [positive] refraction device will yield a negative [positive] k_{e} in our setup, and consistently, opposite [same] signs for $d\omega/dk_{\mathrm{e}}$ and ω/k_{e}. In Fig. 3.12a, $d\omega/dk_{\mathrm{e}}$ and ω/k_{e} have opposite signs in the bright region, and the sign of k_{e} is minus, both of which consistently confirm negative refraction in the bright region. From the dispersion relation in the bright region, the negative refractive index is obtained by $n = k_{\mathrm{e}}c/\omega$.

In the dark region of Fig. 3.12a where the cutoff behavior occurs and the de-embedded $|s_{21}|$ becomes exceedingly small mostly below[8] -30 dB, the extraction method above becomes increasingly error-prone, as discussed in Ref. [59] and as can be seen from Eq. (3.17). This can result in a dispersion relation which is physically not meaningful or mathematically troublesome without being able to satisfy the causality condition [59]. For example, in the dark region of Fig. 3.12a, the group velocity is negative, while the energy actually propagates in the positive reference direction, which is a contradiction; an alternative way of viewing this contradiction is that in the dark region, $k_{\mathrm{e}} < 0$ indicates negative refraction, but the identical signs of $d\omega/dk_{\mathrm{e}}$ and ω/k_{e} suggest positive refraction. Thus, calculating n is meaningless with its sign indeterminable in this cutoff region, and we truncate the effective index curves in cutoff regions. However, we show the dispersion in the dark region in Fig. 3.12a to use the spurious behavior as an indicator of the cutoff region.

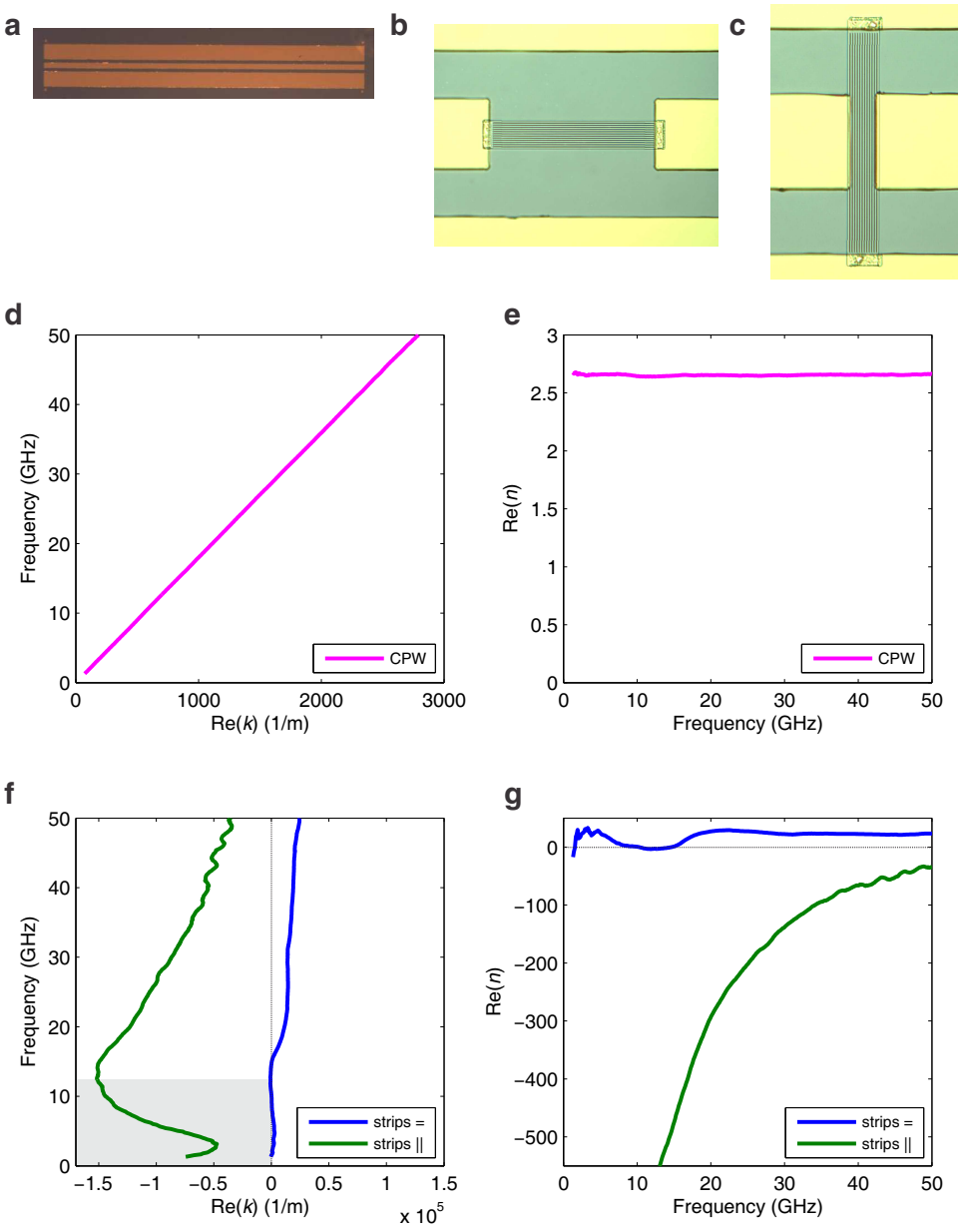

Figure 3.14: Comparison of dispersion relations and effective refractive indices for positively refracting and negatively refracting structures [14]. **a**, 3-mm CPW on GaAs (CPW). **b**, 2DEG strip array excited *along* the direction of the strips (strips =). **c**, The negative index metamaterial device of Fig. 3.12 redisplayed here to facilitate the comparison (strips ||). **d**,**e**, Measured dispersion and effective refractive index, respectively, for the device of **a**. **f**,**g**, Measured dispersions and effective refractive indices, respectively, for the devices of **b** and **c**. All measurements were performed at 4.2 K.

3.2.3.3 Comparison to Positively Refracting Structures

To further illuminate our negative refraction strategy in contrast to positive refraction and also to check the soundness of our methods for calibration, de-embedding, and effective index extraction (albeit well established and broadly used), we have fabricated and measured positively refracting structures. All measurements, calibration, de-embedding (if applicable), and effective index extraction procedures performed on these structures are exactly the same as those used for our negative index metamaterials. As discussed in Sec. 3.2.3.2, because the measured s_{21} and s_{11} parameters set the left-to-right energy propagation direction (*i.e.*, group velocity direction) as the positive reference direction in all of our measurements, if the device under test is positively [negatively] refracting, the extracted k_{e}'s sign will be plus [minus] with no ambiguity, and at the same time, $d\omega/dk_{\mathrm{e}}$ and ω/k_{e} will have the same [opposite] signs. Fig. 3.14 juxtaposes the measurement results of our metamaterial device of Fig. 3.12 with the measurement results of these positively refracting devices.

Fig. 3.14a shows a 3-mm long electromagnetic coplanar waveguide (CPW) on a GaAs substrate. Its measured dispersion relation (Fig. 3.14d) shows the same signs for $d\omega/dk_{\mathrm{e}}$ and ω/k_{e}, and consistently with this, k_{e} is positive, thus, the CPW is positively refracting, as is well known. The effective dielectric constant κ of ~ 7.0 for this CPW is the average of the dielectric constants of GaAs and air. This leads to an effective refractive index of $n = \sqrt{\kappa} \sim 2.6$. This is matched by the measured index (Fig. 3.14e).

Fig. 3.14b shows the same 2DEG strip array as the device measured for Fig. 3.12 ($a = 1.25$ μm, $W = 1.0$ μm, and $l = 112$ μm), but with excitation *along* the strips. We expect positive refraction; in fact, as all strips are excited together, this device does not feel the inter-strip capacitances, and thus is more or less the same as a *sheet* of 2DEG. The measured dispersion indeed shows the positive refraction, consistently judged by $k_{\mathrm{e}} > 0$, and by the same signs of $d\omega/dk_{\mathrm{e}}$ and ω/k_{e} (Fig. 3.14f, blue). The refractive index extracted from measurements is ~ 23 in a good part of the measurement frequency range (Fig. 3.14g, blue), with the lower frequency regions showing fluctuations due to low $|s_{21}|$, where impedance mismatch and losses are severe. The index of this device is larger than that of the electromagnetic CPW, because the former comes from the plasmonic excitation (in the long plasmonic wavelength regime), or equivalently, as it involves collective electron accelerations. The index ~ 23 of this device,

[8] Note that this is smaller than what is shown in Fig. 3.12d, which is the s_{21} magnitude before de-embedding the parasitic signals.

however, is not as large in magnitude as the negative index of our metamaterial, because the inter-strip capacitance does not participate in the dynamics in the former.

Finally, Fig. 3.14c shows our negative index metamaterial measured for Fig. 3.12. Its measured dispersion and index (Figs. 3.12a-b) are repeated here (Figs. 3.14f-g, green) to facilitate the comparison with the positive refraction devices. In Fig. 3.12, the dispersion relation was plotted with $-k_{\mathrm{e}}$ as the x-axis, but here, we use k_{e} as the x-axis; the data are exactly identical (in either case, $k_{\mathrm{e}} < 0$). Both the negative sign of k_{e} and the opposite signs of $d\omega/dk_{\mathrm{e}}$ and ω/k_{e} in the measured dispersion in the bright region consistently confirm negative refraction. Contrast from the positive refraction devices is evident.

3.2.3.4 Pre-De-Embedding Data

Fig. 3.15 corresponds to Figs. 3.12a-c, but without the parasitic coupling de-embedded; as mentioned earlier, Figs. 3.12a-c are obtained after the full de-embedding. By comparing Figs. 3.12a-c and Fig. 3.15, we see that the positively refracting parasitic coupling lowers the magnitude of the measured negative index, as expected. Even when the parasitic coupling is treated as a part of the device response, the measured effective index is still very large on the order of hundreds (Fig. 3.15b). The parasitic coupling is also seen to lower the measured loss and cause the figure of merit to be overestimated (Fig. 3.12c vs. Fig. 3.15c).

Fig. 3.16 shows the entire set of the raw s-parameter data up to the probe tips for the device of Fig. 3.12; previously, only the magnitudes of s_{11} and s_{21} were presented to show the essence of the difference between the 2DEG and parasitic signals. Each set of data, now including phase information, clearly shows that the 2DEG device behaves largely consistently as a negative index metamaterial at temperatures of 4.2 K, 10 K, and 20 K, while its behavior is completely different at 297 K. This once again confirms the negative refraction due to kinetic inductance in the acceleration-dominated regime (4.2 K, 10 K, and 20 K) and the open-circuit behavior in the scattering-dominated regime (297 K). In addition, all of these data show that the 2DEG device's behavior at 297 K much resembles the behavior of the device having only the CPWs with no 2DEG strip array (labeled 'open' in Fig. 3.16); this once again attests to the open-circuit behavior of the metamaterial at 297 K, with the kinetic inductance completely masked by electron scattering.

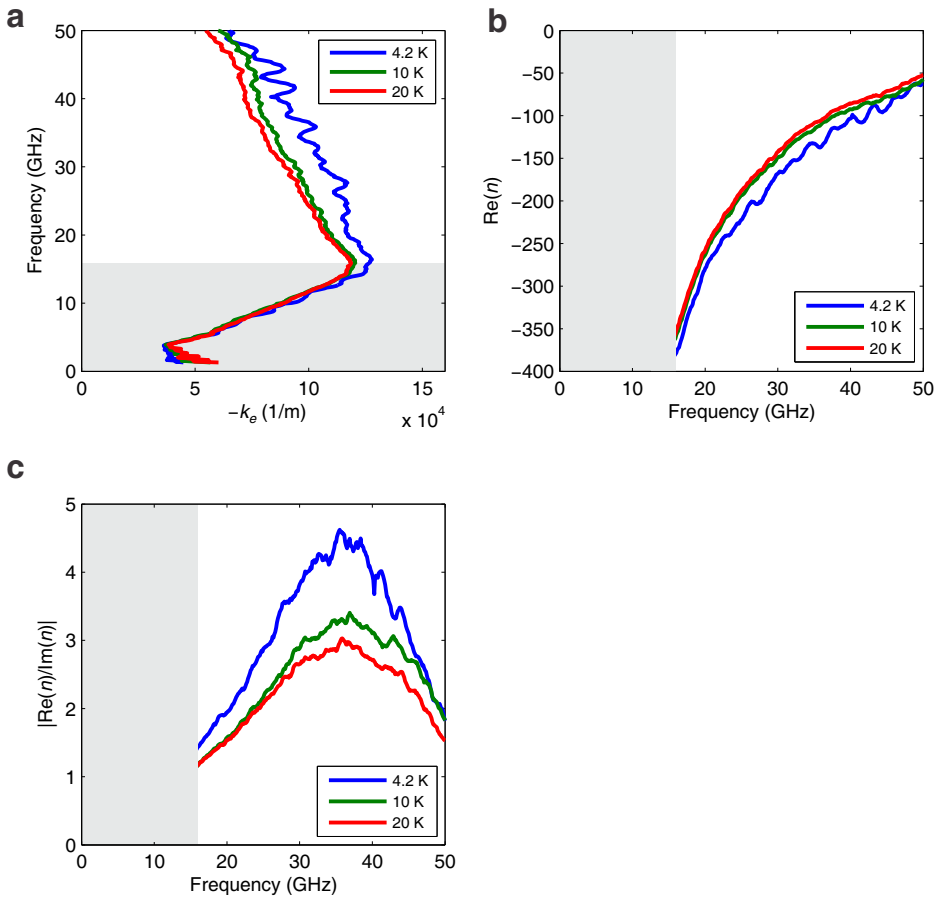

Figure 3.15: Measurement results of the device of Fig. 3.12, without the parasitic coupling de-embedded [14]. Panels **a**, **b**, and **c** correspond to Figs. 3.12a-c.

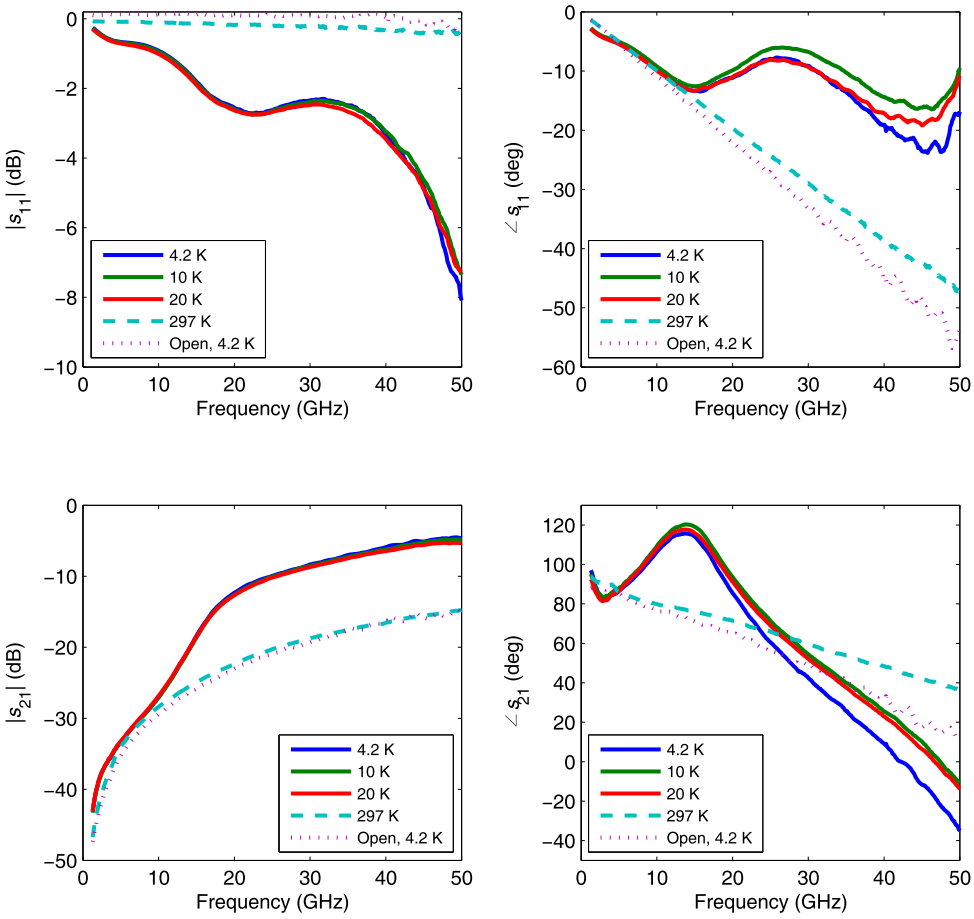

Figure 3.16: Full s-parameter data for the device of Fig. 3.12 [14]. $|s_{11}|$ is slightly (~ 0.1 dB; 2 %) larger than 0 dB for the 'open' device measurement due to the calibration error.

3.2.4 Appendix: Materials and Methods

We fabricate the devices on GaAs/AlGaAs 2DEG substrates obtained by molecular beam epitaxy. The layer structure above the 2DEG is 40-nm $\mathrm{Al_{0.36}Ga_{0.64}As}$, 14-nm Si-doped $\mathrm{Al_{0.36}Ga_{0.64}As}$, 10-nm $\mathrm{Al_{0.36}Ga_{0.64}As}$, and 7-nm GaAs cap. At 4 K, the mobility of the 2DEG is $4.6 \times 10^6\ \mathrm{cm^2/Vs}$, and the carrier density is $1.9 \times 10^{11}/\mathrm{cm^2}$, both in dark. 2DEG strips are defined by electron beam lithography, followed by wet etching ($>$ 71-nm depth) with 150:1:1 $\mathrm{H_2O : H_2O_2 : NH_4OH}$. Ohmic contacts are defined by photolithography followed by thermal evaporation of 5-nm Ni / 20-nm Au / 25-nm Ge / 10-nm Au / 5-nm NI / 40-nm Au, and annealing at $420\ ^\circ\mathrm{C}$ for 50 seconds. CPWs are defined by photolithography and formed by thermal evaporation of 8-nm Cr / 500-nm Au.

The microwave scattering analysis is performed in a Lakeshore cryogenic probe station at feedback-controlled cryogenic temperatures in dark. 100-m pitch ground-signal-ground microwave probes connected to probe arms land on the on-chip CPWs. Coaxial cables lead to the probes from an Agilent E8364A network analyzer, which generates the excitation signals up to 50 GHz with -45 dBm power reaching the devices, and measures the scattering parameters. To see the pure effect of the metamaterial (strip array) only, we first calibrate the system, at each measurement temperature, up to the tips of the probes by using the NIST-style multi-line TRL technique [70], and then perform additional de-embedding to remove the on-chip CPW delays and parasitic couplings between the CPWs, which bypass the metamaterial (see Appendix Ch. A). The CPWs used for this calibration [11] are fabricated on undoped GaAs substrates and designed to have a $50\text{-}\Omega$ characteristic impedance, which is the characteristic impedance of the network analyzer, cables, and probes.

Fig. 3.8b and Table 3.2 were obtained using Sonnet frequency-domain electromagnetic field solver. For a conductor, the simulator takes complex resistivity as a simulation parameter; we use the real part of the resistivity to model the electron scattering effect in the 2DEG, and its imaginary part to model the kinetic inductance of the 2DEG.

The simulated structures lie in a $400\ \mu\mathrm{m} \times 400\ \mu\mathrm{m} \times 1000\ \mu\mathrm{m}$ (x, y, and z) box surrounded by conducting side and bottom walls. Inside the box there are three dielectric layers; from bottom up, the first layer is $500\ \mu$m-thick GaAs with $\kappa_{\mathrm{GaAs}} = 12.9$ (which corresponds to the actual thickness of the GaAs substrate below the 2DEG), the second layer is $71\ \mathrm{nm}$-thick AlGaAs with $\kappa_{\mathrm{AlGaAs}} = 12.9$ (which corresponds to the actual thickness of the AlGaAs and GaAs cap above the 2DEG), and the third layer

is $500\ \mu\mathrm{m}$ thick air with $\kappa_{\mathrm{Air}} = 1.0$. Between the first and second dielectric layers lies the 2DEG structure; the 2DEG's resistance per square $7.2\ \Omega$ and kinetic inductance per square $1.25\ \mathrm{nH}$ are entered into the aforementioned complex resistivity parameter. Between the second and third dielectric layers, we define the gold CPWs with conductivity $4.09 \times 10^7\ \mathrm{S/m}$ and thickness $0.5\ \mu\mathrm{m}$ (which corresponds to the actual thickness of the CPWs). In the actual structure, most parts of the CPWs sit directly on top of GaAs, instead of AlGaAs as defined in the simulation, as most of AlGaAs except where the 2DEG strips are defined is etched away. However, this makes only a negligible difference as the 71-nm thickness of the AlGaAs layer is much smaller than the relevant wavelengths.

To model the strip array of Fig. 3.12, for instance, the 2DEG is formed into strips of width $1\ \mu\mathrm{m}$ and length $112\ \mu\mathrm{m}$, separated from each other with $0.25\ \mu\mathrm{m}$ gaps (*i.e.*, periodicity is $1.25\ \mu\mathrm{m}$). There are a total of 13 strips in the strip array. The ohmic contact between each 2DEG strip and the CPWs' ground is modeled as a $1\ \mu\mathrm{m} \times 1\ \mu\mathrm{m}$ conductor (at the 2DEG layer level) with a resistance of $450\ \Omega/\square$, which is an estimation (the real contact resistance varies from device to device); the ohmic contact is completed by inserting a resistance-less vertical via to the gold CPW layer.

CPWs are also designed using Sonnet to possess a 50-Ω characteristic impedance on air-clad GaAs substrate. The dimensions of the CPWs are signal line width $50\ \mu\mathrm{m}$, gaps between the signal line and the ground lines $32\ \mu\mathrm{m}$, and ground line width $143\ \mu\mathrm{m}$ (top and bottom). In the real devices, slightly narrower ($120\ \mu\mathrm{m}$) ground lines were used due to space constraints on the small 2DEG samples, but this causes only a negligible change in the characteristic impedance of the CPWs, and the measured signal as well, as confirmed by additional simulations and experiments. The signal lines extend over the first two 2DEG strips from both sides (left and right) of the strip array.

Chapter 4

2D Plasmonics in Graphene

4.1 Measurement of Collective Mass of Graphene Electrons

Individual electrons in graphene[1] behave as massless quasiparticles [3, 26, 71–75]. In surprising twist, it is inferred from plasmonic investigations [12, 27, 76, 77] that collectively excited graphene electrons must exhibit non-zero mass and its inertial acceleration is essential for graphene plasmonics. Despite such importance, this collective mass has defied direct unequivocal measurement. It may be directly measured by accelerating it with a time-varying voltage and quantifying the phase delay of the resulting current; this voltage-current phase relation would manifest as kinetic inductance, representing the collective inertia's reluctance to accelerate. However, at optical (infrared) frequencies phase measurement of current is generally difficult and at microwave frequencies the inertial phase delay has been buried under electron scattering [78–80]. Here we directly, precisely measure the kinetic inductance, thus, collective mass, by combining innovative device engineering that reduces electron scattering and delicate microwave phase measurements. Particularly, encapsulation of graphene between hexagonal-boron-nitride layers [67], one-dimensional edge contacts [81], and a proximate top gate configured as microwave ground [11,13] together enable resolving the inertial phase delay from the electron scattering. Beside the fundamental importance, the kinetic inductance demonstrated here to be orders-of-magnitude larger than magnetic inductance can dramatically miniaturize radio-frequency integrated circuits. Moreover, its bias-dependency heralds a solid-state voltage-controlled inductor to complement the prevalent

[1] Large portions of this section are derived from a paper in publication by the author [16].

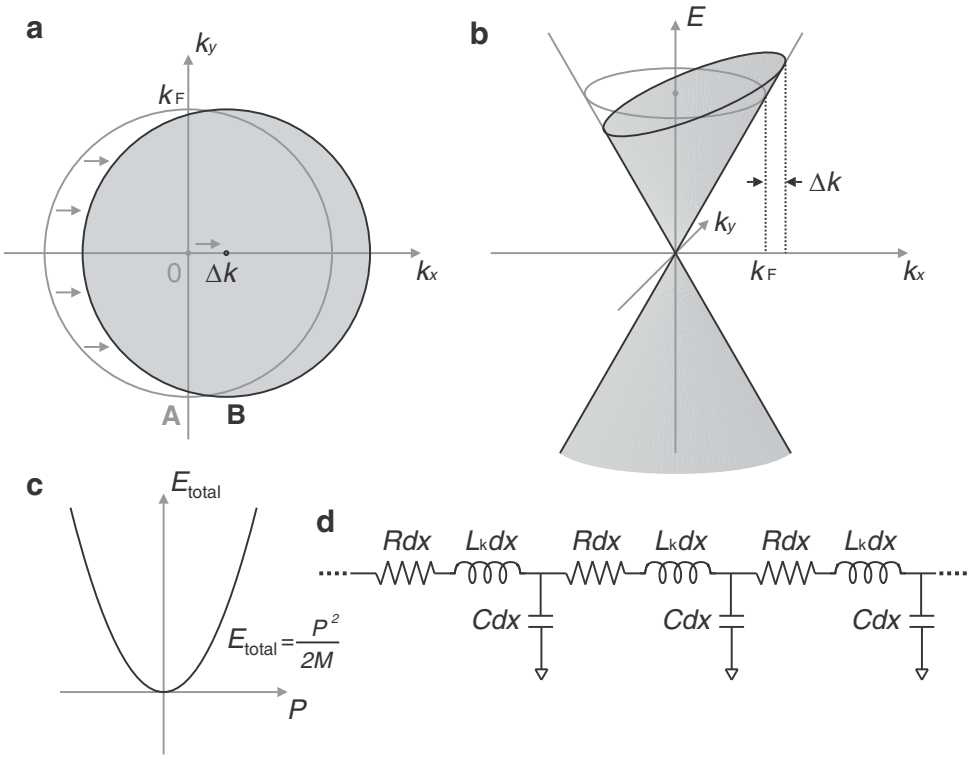

Figure 4.1: Collective electrodynamics of graphene electrons [16]. **a**, Collective motion of graphene electrons subjected to an electric field can be represented as a translation of Fermi disk in k-space. **b**, Alternative representation of the collective electron motion in the E-k space, in conjunction with the massless single electron energy dispersion $E = \hbar v_{\mathrm{F}} k$ near the Dirac point. **c**, The per-unit-length collective kinetic energy E_{total} exhibits quadratic dependency on the per-unit-length collective momentum $P = n_0 W \hbar \Delta k$. The curvature is inversely proportional to the per-unit-length collective dynamical mass, M. **d**, Graphene as a lossy transmission line.

voltage-controlled capacitor.

4.1.1 Collective Mass of Graphene Electrons

4.1.1.1 Concept of Collective Mass

The collective excitation of massless fermions in graphene exhibits a non-zero mass. This fact is subsumed under the general theoretical framework of graphene plasmonics [27], yet it can be simply seen as follows. Let electrons in graphene (width W, unit length) be subjected to an electric field along the length with a voltage difference V across the length. The resulting translation of the Fermi disk in k-space by $\Delta k \ll k_{\mathrm{F}}$ (from disk **A** to **B**, Figs. 4.1a,b) yields a per-unit-length collective translational wavenumber, $K = n_0 W \Delta k$, and a per-unit-length collective momentum, $P = \hbar K = n_0 W \hbar \Delta k$. The corresponding

per-unit-length collective kinetic energy[2] $E_{\rm total}$ is obtained by subtracting the sum of single electron energies $E = \hbar v_{\rm F} k$ for disk **A** from that for disk **B**. Since $E_{\rm total}$ assumes an extremum (minimum) when $\Delta k = 0$, *i.e.*, in the collective ground state, we must have $E_{\rm total} \propto (\Delta k)^2 \propto P^2$ for small Δk (Fig. 4.1c). In fact, calculation to the lowest order of Δk shows

$$E_{\rm total} = \frac{E_{\rm F} W}{2\pi} (\Delta k)^2 = \frac{1}{2} \frac{E_{\rm F}}{\pi n_0^2 W} K^2 = \frac{1}{2} \frac{E_{\rm F}}{\pi \hbar^2 n_0^2 W} P^2, \tag{4.1}$$

which is identical to Eq. (2.14) with $g = 4$ (spin and valley degeneracy in graphene), but rewritten here in terms of the collective wavenumber K and collective momentum P.

Hence, while individual electrons with the linear E-k dispersion ($E = \hbar v_{\rm F} k$) behave as massless particles, electrons in collective dynamics with the quadratic relation $E_{\rm total} \propto (\Delta k)^2 \propto K^2 \propto P^2$ must exhibit a non-zero collective mass per unit length

$$M = \frac{\pi \hbar^2 n_0^2 W}{E_{\rm F}}, \tag{4.2}$$

with which Eq. (4.1) can be written as

$$E_{\rm total} = \frac{\hbar^2 K^2}{2M} = \frac{P^2}{2M}. \tag{4.3}$$

The velocity $v_{\rm c}$ of the collection of electrons is then given by

$$v_{\rm c} = \frac{1}{\hbar} \frac{\partial E_{\rm total}}{\partial K} = \frac{\hbar K}{M} = \frac{P}{M}. \tag{4.4}$$

As K increases with time t by per-unit-length force F (*e.g.*, due to the electric field considered here) applied to the unit length of the collection of electrons in the graphene strip, $v_{\rm c}$ increases with t. Since $E_{\rm total} = \int_{t_0}^{t} F v_{\rm c} dt = P^2/(2M)$ (t_0 is when the electron gas is at rest), by differentiating this relation with respect to t, we obtain

$$\frac{d}{dt}(M v_{\rm c}) = F. \tag{4.5}$$

That is, the collective mass of electrons in graphene 'inertially' accelerates according to the Newton's

[2] Unlike in Chapter 2, we here express the collective kinetic energy *per unit length* for convenience in notations.

2nd law. Note the contrast between the collective velocity v_{c} and the individual electron velocity given by $\frac{1}{\hbar}\frac{\partial E}{\partial k} = v_{\mathrm{F}}$, which is constant even in the presence of the force F.

In sum, in graphene, while an individual electron behaves like a massless particle with a constant velocity v_{F}, electrons in collective excitation together exhibit a non-zero collective inertial mass and a variable collective velocity v_{c} in keeping with the Newton's equation of motion. The emergence of the non-zero collective electron mass despite the zero individual electron mass in graphene sharply contrasts typical conductors where the non-zero individual electron effective mass m^* appearing in the quadratic single electron energy dispersion $E = \hbar^2 k^2/(2m^*)$ is plainly multiplied by the total number of electrons to give the collective mass. For instance, consider a unit-length strip (width W) of GaAs/AlGaAs 2DEG. We can repeat the calculation above but with $E = \hbar^2 k^2/(2m^*)$ and $g = 2$ to obtain $E_{\mathrm{total}} = \frac{\hbar^2 k_{\mathrm{F}}^2 W}{4\pi m^*}(\Delta k)^2$. By using $k_{\mathrm{F}}^2 = 4\pi n_0/g$ and $P = n_0 W \hbar \Delta k$, we can see $E_{\mathrm{total}} = \frac{1}{2}\frac{1}{W n_0 m^*}P^2$, *i.e.*, the collective mass per unit length is m^* multiplied by the total number of electrons contained in the unit-length GaAs/AlGaAs 2DEG.

To estimate how large the collective mass of electrons in graphene is, we can *operationally* calculate the collective mass *per electron*: $m_{\mathrm{c}}^* = M/(W n_0) = E_{\mathrm{F}}/v_{\mathrm{F}}^2$. For $E_{\mathrm{F}} = 0.1$ eV, $m_{\mathrm{c}}^* = 0.02 m_{\mathrm{e}}$. For comparison, in GaAs/AlGaAs 2DEG, $m^* = 0.067 m_{\mathrm{e}}$. Incidentally, we note that this collective mass per electron is quantitatively related to an insightfully defined theoretical entity called 'plasmon mass' in graphene [76,77,82]; the former, which we set out to measure here, is an observable physical reality that proves the existence of the latter beyond a theoretical model.

The collective current I associated with the Fermi disk shift, *i.e.*, the inertial acceleration of the collective mass M, has an inductive phase relationship to the applied voltage V that causes the acceleration, where the associated inductance is kinetic inductance. The kinetic inductance can be evaluated by noting $E_{\mathrm{total}} \propto I^2$, given $I \propto \Delta k$ for small Δk and $E_{\mathrm{total}} \propto (\Delta k)^2$; by analogy to magnetic inductance, this energy can be then expressed as $E_{\mathrm{total}} = \frac{1}{2} L_{\mathrm{k}} I^2$, where

$$L_{\mathrm{k}} = \frac{\pi \hbar^2}{e^2 E_{\mathrm{F}}}\frac{1}{W} = \frac{\hbar\sqrt{\pi}}{e^2 v_{\mathrm{F}}\sqrt{n_0}}\frac{1}{W} \tag{4.6}$$

is the per-unit-length kinetic inductance, which is identical to Eq. (2.18). The same underlying physics, namely quadratic dependence of E_{total} on Δk, gives rise to both M and L_{k}, which are thus intimately

related by

$$M = e^2 n_0^2 W^2 L_{\rm k}. \tag{4.7}$$

Typical conductors with the quadratic single electron E-k dispersion also exhibit kinetic inductance arising from their more obvious collective electron mass, and $L_{\rm k} \propto 1/n_0$ [11, 13, 20]. By contrast, in graphene, $L_{\rm k} \propto 1/\sqrt{n_0}$. This peculiarity of graphene results from the linear single electron E-k dispersion; that is, while individual and collective behaviors of electrons in graphene exhibit vast difference with zero individual electron mass and non-zero collective electron mass, the values of the collective electron mass and kinetic inductance are directly affected by the single electron energy dispersion.

To weigh M, we directly measure $L_{\rm k}$ essentially by interrogating the voltage-current phase relation in graphene via microwave transport experiment, where graphene acts as a lossy transmission line [11, 13,14,20,22] (Fig. 4.1d). This entails the per-unit-length kinetic inductance $L_{\rm k}$ modeling local collective mass and per-unit-length geometric capacitance C terminated to ground specific to device configuration. The magnetic inductance, which is orders of magnitude smaller [11, 14] than $L_{\rm k}$ expected in graphene (see Sec. 4.1.2.3), and quantum capacitance [19], whose effect is far weaker than that of geometric capacitance in our device geometry to be discussed, are both ignored. The per-unit-length resistance R models electron scattering.

While the graphene kinetic inductance can be implied from the plasmonic theory [27] and has been considered in explicit theories [15,21,82], its direct measurement has been evasive. In far-infrared intensity transmission spectroscopies, kinetic inductance can be indirectly inferred from the fitting parameter called Drude weight [76, 77, 82], but as these experiments do not measure the phase progression of the collective current, they do not unambiguously prove the existence of the collective mass and its inertial acceleration. At microwave frequencies, while $L_{\rm k}$ can in principle be directly measured by probing the voltage-current phase relation, experimental attempts [78–80] have proven unfruitful because R is far larger than the inductive impedance $i\omega L_{\rm k}$ at microwave frequencies (*i.e.*, the kinetic inductor's quality factor $Q = \omega L_{\rm k}/R$ is far smaller than 1) even in reasonably high-mobility graphene, although $L_{\rm k}$ is far larger than magnetic inductance.

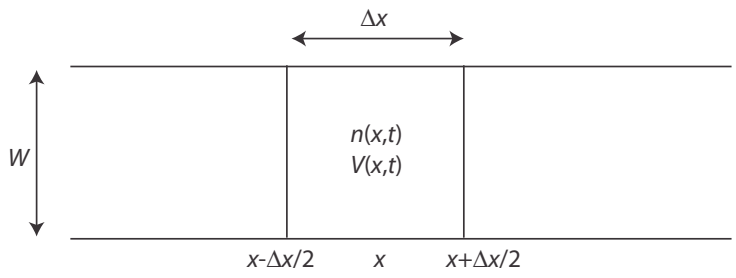

Figure 4.2: Electron density and potential distribution along a graphene strip where a plasmonic wave propagates along the length (x-axis) [16]. The infinitesimal length Δx is exaggerated.

4.1.1.2 Collective Mass and Plasmonic Wave

In Chapter 2, we saw how a 2D plasmonic wave can propagate on a 2D conductor in terms of $L_{\rm k}$ and C, leading to a plasmonic dispersion relation $v_{\rm p} = \omega/k_{\rm p} = 1/\sqrt{L_{\rm k}C}$. Instead, we can visualize the link between the graphene plasmonic wave and the collective electron mass without using the circuit (transmission line) model, but by directly considering the dynamics of the collective electron mass in the presence of Coulomb restoring force and electron-degeneracy restoring pressure.

Imagine a plasmonic wave propagating along the length (x-axis) of the graphene strip. Let the potential and electron density of the infinitesimal segment of length Δx at position x be $V(x,t)$ and $n(x,t) = n_0 + \delta_n(x,t)$, respectively (Fig. 4.2)[3], where $\delta_n(x,t)$ accounts for the extra (excess or deficit) electron density in this infinitesimal segment. The extra charge due to the extra electron density and potential $V(x)$ of the segment are related through its capacitance, $C\Delta x$, where $C = C_{\rm c}C_{\rm q}/(C_{\rm c} + C_{\rm q})$:

$$-eW\Delta x\delta_n(x,t) = C\Delta x V(x,t) \tag{4.8}$$

Since the restoring force per unit positive charge is $-\frac{\partial}{\partial x}V(x,t)$, the restoring force that collectively drives the collection of electrons in the infinitesimal segment is given by

$$F = -[-en(x,t)W\Delta x]\frac{\partial}{\partial x}V(x,t) = -\frac{e^2 n(x,t)W^2\Delta x}{C}\frac{\partial}{\partial x}\delta_n(x,t), \tag{4.9}$$

where we have used Eq. (4.8) in obtaining the last expression. With this force expression at hand, we

[3]The potential $V(x,t)$ is due not only to the Coulomb force, but also to the electron degeneracy pressure, that is, it combines the electric and quantum-mechanical potential. We include this quantum effect here for generality, although its effect in our actual graphene device is negligible.

now set up the equation of motion for the collection of electrons in the infinitesimal segment according to Eq. (4.5) as

$$\frac{\partial}{\partial t}[M(x,t)\Delta x v_{\rm c}(x,t)] = -\frac{e^2 n(x,t) W^2 \Delta x}{C}\frac{\partial}{\partial x}\delta_n(x,t). \tag{4.10}$$

Here $v_{\rm c}(x,t)$ is the collective velocity of electrons in the infinitesimal segment and $M(x,t)$ is the per-unit-length collective electron mass at position x (thus $M(x,t)\Delta x$ is the collective electron mass of the infinitesimal segment). Given Eq. (4.2), $M(x,t)$ can be expressed as

$$M(x,t) = \frac{\pi\hbar^2 n^2(x,t) W}{E_{\rm F}(x,t)}, \tag{4.11}$$

where $E_{\rm F}(x,t)$ is the Fermi energy corresponding to $n(x,t)$. Since $\delta_n(x,t) \ll n_0$ in any practical situation, in the two equations above, we may use $n(x,t) \approx n_0$ and $E_{\rm F}(x,t) \approx E_{\rm F}$, where $E_{\rm F}$ is the previously defined equilibrium Fermi energy corresponding to n_0. Then Eq. (4.10) reduces to

$$\frac{\partial}{\partial t} v_{\rm c}(x,t) = -\frac{1}{L_{\rm k} C n_0}\frac{\partial}{\partial x}\delta_n(x,t). \tag{4.12}$$

where we have used Eq. (4.6). On the other hand, as the total number of electrons is conserved, $v_{\rm c}(x,t)$ and $n(x,t)$ satisfy

$$\frac{\partial}{\partial x}[n(x,t) v_{\rm c}(x,t)] = -\frac{\partial}{\partial t} n(x,t). \tag{4.13}$$

By using $n(x,t) = n_0 + \delta_n(x,t)$ on both sides, but noting that $n(x,t)$ can be approximated as $n(x,t) \approx n_0$ on the left hand side, we obtain

$$n_0 \frac{\partial}{\partial x} v_{\rm c}(x,t) = -\frac{\partial}{\partial t}\delta_n(x,t). \tag{4.14}$$

By combining Eqs. (4.12) and (4.14), plasmonic wave equations are obtained as

$$\begin{aligned} \frac{\partial^2}{\partial x^2}\delta_n(x,t) &= L_{\rm k} C \frac{\partial^2}{\partial t^2}\delta_n(x,t), \\ \frac{\partial^2}{\partial x^2} v_{\rm c}(x,t) &= L_{\rm k} C \frac{\partial^2}{\partial t^2} v_{\rm c}(x,t). \end{aligned} \tag{4.15}$$

These plasmonic wave equations confirm $v_{\rm p} = \omega/k_{\rm p} = 1/\sqrt{L_{\rm k} C}$, which is the plasmonic dispersion

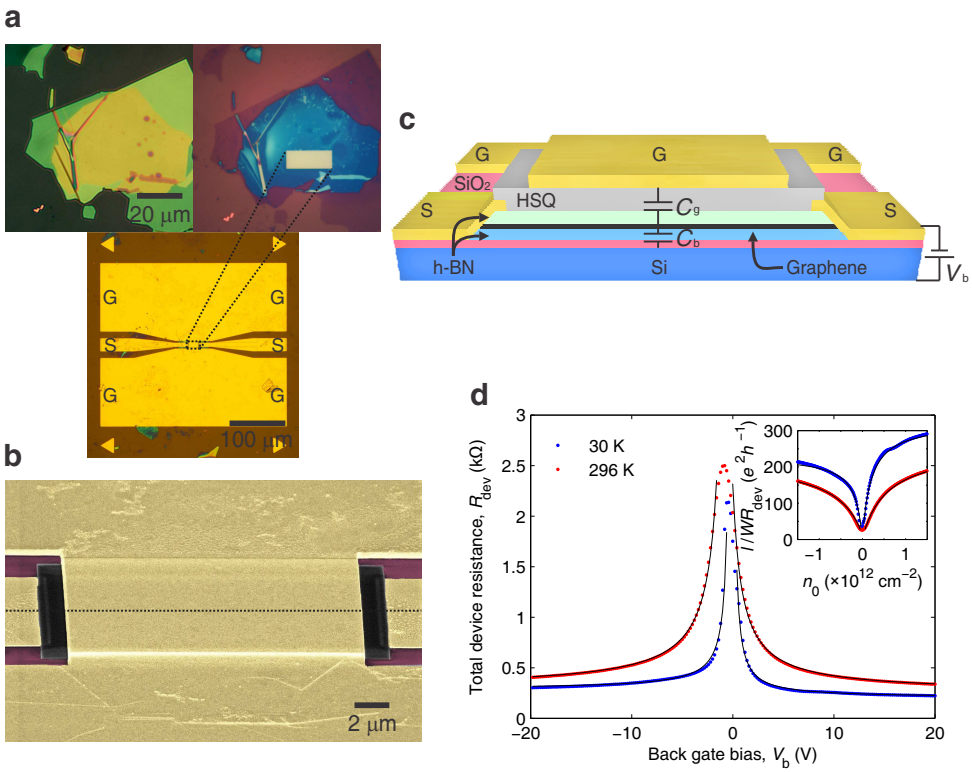

Figure 4.3: Device description and DC measurements [16]. **a**, Optical image of the h-BN/graphene/h-BN layered structure before etching (top-left), after etching (top-right), and after depositing the CPWs (bottom). **b**, False colored scanning electron micrograph of the central region of the device that contains the layered structure under the top gate. **c**, Schematic diagram of h-BN encapsulated graphene device with the front face corresponding to the vertical cut through the dotted line in **b**. **d**, Total device resistance $R_{\rm dev}$, including both in-graphene electron scattering effect R and contact resistance, measured at 30 K and 296 K with $V_{\rm b}$ varied while graphene and the top gate are kept at the same DC potential. (inset: corresponding plot of $(R_{\rm dev}/(l/W))^{-1}$, a conductivity estimate including contact effects; $n_0 = C_{\rm b}/W \times (V_{\rm b} - V_{\rm b,0})/e$ with $C_{\rm b}/W = 0.12$ fF/μm^2 and $V_{\rm b,0} = -0.5$ V (30 K) or $V_{\rm b,0} = -0.9$ V (296 K)). Solid curves are fits to $\sigma^{-1} = (n_0 e \mu_{\rm C})^{-1} + \rho_{\rm s}$.

relation previously obtained in Chapter 2. This dynamic consideration once again attests to the physical reality of the collective electron mass, by directly delineating the machinery of the participation of the collective mass in the plasmonic wave propagation.

4.1.2 Device Measurements Summary

In our microwave measurements of $L_{\rm k}$, we help overcome the difficulty faced by previous works [78–80] by reducing electron scattering, thus, R, as much as possible. In particular, we encapsulate exfoliated graphene between two hexagonal boron nitride (h-BN) layers (Fig. 4.3) by a polymer-free assembly method [67, 81], which greatly reduces electron scattering by disorder. To reduce electron-phonon scat-

tering within graphene, the device is cooled to 30 K. Furthermore, to reduce additional electron scattering in the contact regions at both ends of graphene, we make one-dimensional, edge-only contacts to the graphene by etching the stack of h-BN and graphene into a desired shape ($W = 7.5\ \mu\mathrm{m}$; length $l = 19.0\ \mu\mathrm{m}$) and depositing metal onto the side edges [81] (Figs. 4.3a-c). Graphene is connected to the signal (S) lines of on-chip coplanar electromagnetic waveguides (CPWs) to the left and right via the aforementioned edge-only contacts, and is placed under a top gate merged with the ground (G) lines of the CPWs.

4.1.2.1 DC Resistance Measurements

We first measure the DC resistance between the S lines using a lock-in technique, with the graphene and top gate kept at the same DC potential. The total device resistance R_{dev}, including the effect of both R and contact resistance, is measured as a function of the back gate potential V_{b} that sets the carrier density n_0 (Fig. 4.3d). At 30 K, it shows a charge neutrality at $V_{\mathrm{b},0} = -0.5$ V and excellent performance in the electron-doped region ($V_{\mathrm{b}} > V_{\mathrm{b},0}$), which is fit well by the widely-adopted conductivity formula [67, 83] $\sigma^{-1} = (n_0 e \mu_{\mathrm{C}})^{-1} + \rho_{\mathrm{s}}$, with $\mu_{\mathrm{C}} = 390,000\ \mathrm{cm}^2/\mathrm{Vs}$ representing the n_0-independent mobility due to long-range scattering, and $\rho_{\mathrm{s}} = 80\ \Omega$ representing the short range scattering. We note that this is only a lower bound of the actual graphene mobility, because the estimation of conductivity in this two-probe measurement includes the contact effects. The hole-doped region ($V_{\mathrm{b}} < V_{\mathrm{b},0}$) shows similar results, but with slightly lower $\mu_{\mathrm{C}} = 320,000\ \mathrm{cm}^2/\mathrm{Vs}$ and higher $\rho_{\mathrm{s}} = 110\ \Omega$ due to the contact characteristics [81]. The room-temperature measurement shown for comparison (which exhibits a slightly shifted neutrality at $V_{\mathrm{b},0} = -0.9$ V) has a 4 times smaller μ_{C} of $110,000\ \mathrm{cm}^2/\mathrm{Vs}$ in the electron-doped regime, which still is an excellent number. This high μ_{C} at room temperature confirms the reduction in electron scattering by the h-BN encapsulation [67, 81], and its 4-fold increase at 30 K confirms the further scattering reduction at the low temperature.

4.1.2.2 Microwave Measurements and L_{k} Extraction

To measure L_{k}, a vector network analyzer launches microwaves (10-50 GHz) onto the CPWs, and records the amplitude and phase response of their transmission (s_{21}, s_{12}) and reflection (s_{11}, s_{22}) (Fig. 4.4a). The network analyzer connects to graphene via cables, probes, and the CPWs, whose phase delay and loss

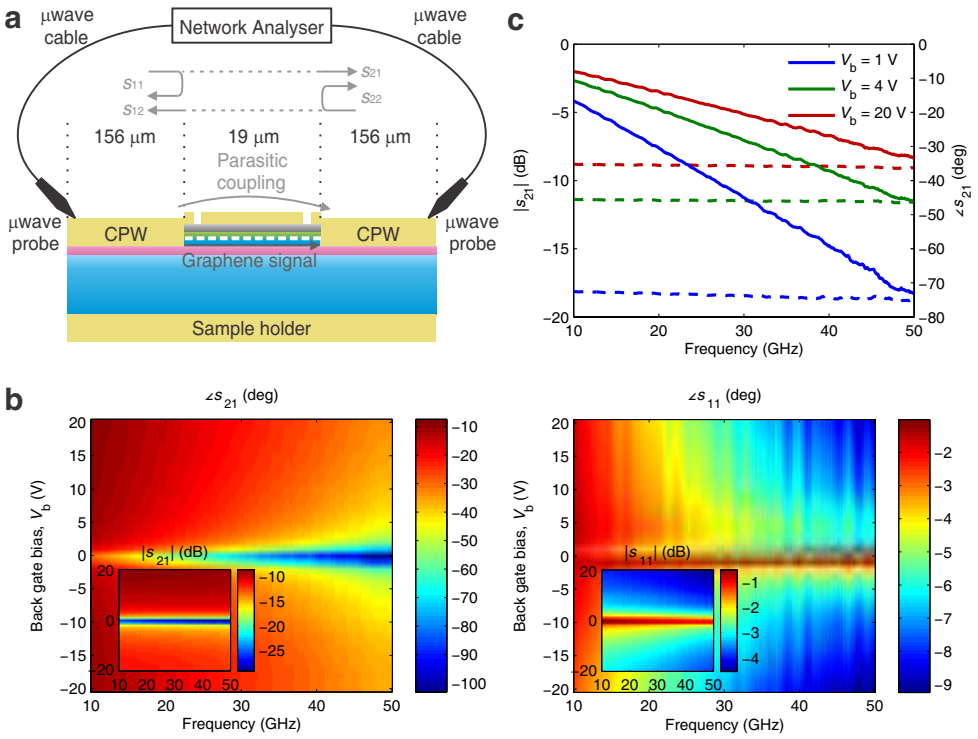

Figure 4.4: Microwave s-parameter measurements [16]. **a**, Schematic diagram of the measurement setup. The s-parameters shown are after calibrating out the delay and loss of the cables, probes, and on-chip CPWs, and also after de-embedding the parasitic coupling bypassing graphene. **b**, Phase (insets: amplitude) of the measured transmission (s_{21}; left) and reflection (s_{11}; right) parameters after the calibration and de-embedding at 30 K. The s-parameters with excitation from the opposite side (s_{12} and s_{22}; not shown here) look almost identical to s_{21} and s_{11}. **c**, Select data from **b**, specifically, transmission phase ($\angle s_{21}$; solid curves) and amplitude ($|s_{21}|$; dashed curves) at three representative bias values $V_\mathrm{b} = 1$, 4, and 20 V.

are calibrated out. The direct parasitic coupling between the left and right CPWs/probes bypassing the graphene channel is separately measured and de-embedded (see Appendix A for details on calibration and de-embedding procedure). We extract L_k from the resulting s-parameters. This extraction, however, poses a stiff challenge for moderate-mobility graphene with $Q \ll 1$, which can be appreciated from the expression for the per-unit-length phase delay ϕ through the graphene transmission line (Fig. 4.1d),

$$\phi \approx \sqrt{\frac{\omega RC}{2}} + \sqrt{\frac{\omega^3}{8}}\sqrt{\frac{C}{R}}L_\mathrm{k} \equiv \phi_1 + \phi_2, \tag{4.16}$$

with only ϕ_2 containing L_k (see Sec. 4.1.3.1 for the derivation), capturing the phase delay due to the collective mass acceleration. The ratio $\phi_2/\phi_1 = \omega L_\mathrm{k}/2R = Q/2$; with $Q \ll 1$, extraction of L_k is challenging because ϕ_2 is entirely swamped by ϕ_1, which typically renders ϕ_2 itself minuscule below the unavoidable phase measurement error, which we denote as ϕ_e, caused by imperfect calibration and non-ideal parasitic signal de-embedding.

To enable L_k extraction from the measured s-parameters, we first reduce R via the aforementioned h-BN encapsulation of graphene and 30-K operation, which amplifies ϕ_2 and attenuates ϕ_1 with improved $Q = 2\phi_2/\phi_1$. This crucial improvement alone, however, is insufficient with the improved Q still smaller than 1. A second improvement is to enhance C by proximate top gating. Although increasing C does not change $Q = 2\phi_2/\phi_1$, it further increases ϕ_2 to ensure $\phi_2 > \phi_\mathrm{e}$. Importantly, these improvements also make ϕ_2 more sensitive to L_k variation, as seen from the factor $\sqrt{C/R}$ in ϕ_2, thereby increasing the accuracy of L_k extraction (see Sec. 4.1.3.1 for an in-depth discussion). To estimate the enhanced value of C in our device, we note that the proximate top gate merged with the CPWs' G lines (Figs. 4.3a-c) serves as a well-defined microwave ground [11, 13] with per-unit-length capacitance C_g between graphene and this grounded top gate. In contrast, the silicon back gate untapped to the G lines 'floats' in microwave signaling, largely because its connection to the DC bias line exhibits a very large inductive impedance and also because the silicon has a high resistivity. Therefore, the per-unit-length capacitance C_b between graphene and the back gate is irrelevant for microwave signaling, and $C = C_\mathrm{g}$. As 44-nm thick top h-BN ($\kappa \approx 7$ [84, 85]) and ∼150-nm thick hydrogen silsesquioxane (HSQ; $\kappa \approx 2.8 \sim 3.0$ [86]) lie between graphene and the grounded top gate, C_g/W is estimated to be $0.15\ \mathrm{fF}/\mu\mathrm{m}^2$, which is far larger than the capacitance of ungated graphene [22]. Incidentally, we attribute the inability to observe [78, 79] L_k or

its spurious measurement [80] in prior works to their larger R with graphene on SiO_2 and no proximate gate configured as a microwave ground.

Figure 4.4b is a color map of the phase and amplitude of transmission (s_{21}) and reflection (s_{11}) parameters measured at 30 K as functions of V_b (thus n_0) and frequency. The s_{21} [s_{11}] amplitude exhibits a sharp drop [peak] near $V_{b,0} = -0.5$ V. Fig. 4.4c shows s_{21} at three select V_b values in the electron-doped region to show that our device s-parameters are amenable to L_k extraction. Were it not for the R-reduction and C-enhancement, the measured s_{21} phase and its portion contributed by L_k, which are intimately related to ϕ and ϕ_2, respectively, would exhibit far smaller absolute values as well as far smaller differences with the variation of V_b (thus with the corresponding variation of L_k), hampering L_k extraction (see Sec. 4.1.3.2 for a detailed discussion).

To determine L_k from the measured s-parameters for each bias, we use the microwave optimization method [11]; we add contact models to both sides of the transmission line model (Fig. 4.1d with $C = C_g$), and alter the component values (e.g., L_k, C_g, R, and contact resistance) until the s-parameters calculated from the model best fit the measured s-parameters across the frequency range in the least-square method (Sec. 4.1.3.3 discusses this procedure in detail). In this way, we determine L_k and other component values at each V_b. This method's reliability is based on the model's physicality and the fact that the limited number of model components must reproduce the vastly larger number of measured s-parameters over the frequency range. Its cogency will be checked ultimately by the consistency amongst the extracted values and other measured parameters, and with the physical theory. The same experiment repeated on a completely different device led to almost identical results (see Sec. 4.1.2.4), further attesting to the reliability of this approach.

Figures 4.5a-c display L_k, C_g, and R so determined for each V_b at 30 K and 296 K. We first focus on the 30-K results in the electron-doped region ($V_b > V_{b,0}$) that showed the best DC characteristics (Fig. 4.3d), in particular in the region away from $V_{b,0}$ (unshaded region in Fig. 4.5). As expected, the extracted C_g/W stays nearly constant with negligible variation from quantum capacitance effect [19], and its value of ~ 0.15 fF/μm^2 is consistent with the value roughly estimated earlier. Also, the extracted L_k closely follows the theoretical curve given by Eq. (4.6) with $v_F = 10^6$ m/s. The slight discrepancy between the observed and theoretical L_k in this region is attributed dominantly to imperfect calibration and parasitic-signal de-embedding, but also potentially to variations of v_F due to dielectric

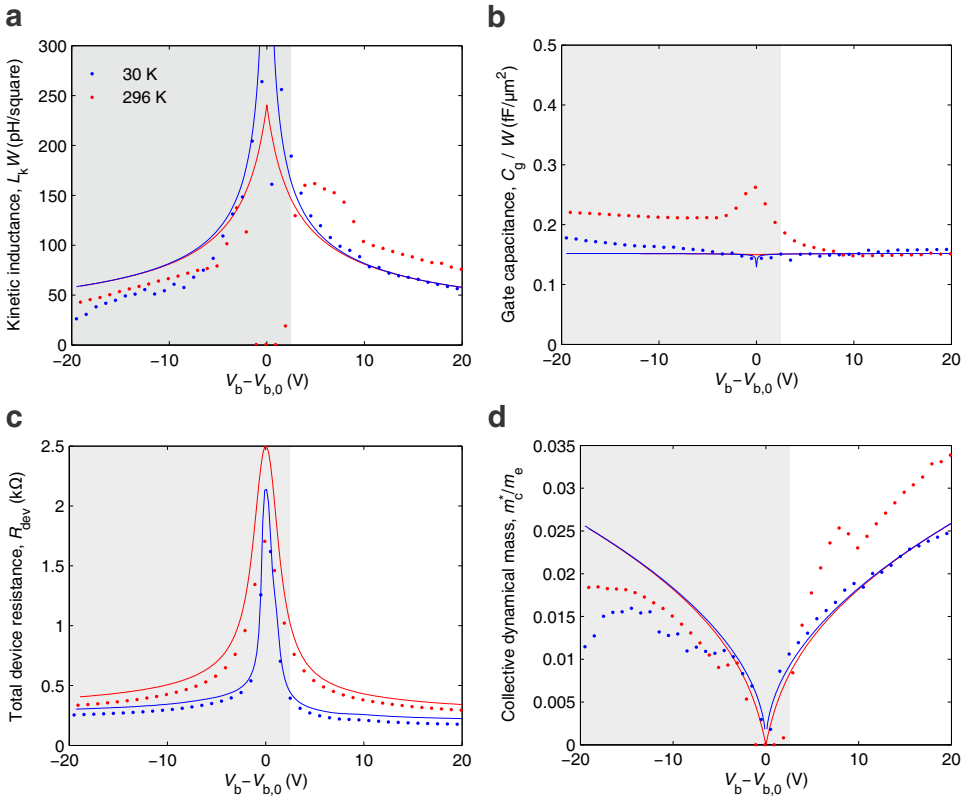

Figure 4.5: Extracted graphene kinetic inductance and collective electron mass [16]. Kinetic inductance per square, $L_{\rm k}W$ (**a**), graphene to top-gate capacitance per unit area, $C_{\rm g}/W$ (**b**), total device resistance, $R_{\rm dev}$ (**c**), and collective dynamical mass per electron, $m_{\rm c}^*$ (**d**), extracted from the measured s-parameters for various $V_{\rm b}$ at 30 K and 296 K. The solid curves in **a**, **b**, and **d** represent theoretical predictions. The solid curve in **c** is $R_{\rm dev}$ measured at DC (Fig. 4.3d). The shaded areas indicate bias regions where the extraction was less reliable (see text).

screening and impurities [85], and/or electron-electron interaction effects [82]. Further confirming the consistency of the technique, $R_{\rm dev}$ extracted from the s-parameters agrees well with $R_{\rm dev}$ measured at DC (Fig. 4.5c). Most importantly, from the measured $L_{\rm k}$, we obtain the per-unit-length collective mass, $M = e^2 n_0^2 W^2 L_{\rm k}$, or operationally defined collective mass per electron that Ref. [82] theorizes as 'plasmon mass,' $m_{\rm c}^* = M/(W n_0) = e^2 n_0 W L_{\rm k}$, which closely follows the theoretical prediction (Fig. 4.5a, inset); $m_{\rm c}^*$ is a few percent of $m_{\rm e} = 9.1 \times 10^{-31}$ kg.

Near the charge neutrality point or in the hole-doped region ($V_{\rm b} < V_{\rm b,0}$) (shaded region, Fig. 4.5), the extracted values of $L_{\rm k}$, C, and $m_{\rm c}^*$ at 30 K exhibit more appreciable deviation from theory. The discrepancy near the charge neutrality is readily understood, because transmission amplitude is significantly smaller due to the sharply reduced n_0 (Fig. 4.4b). In this region, the raw transmission s-parameters before removing the graphene-bypassing parasitic signal are dominated by the parasitic signal itself, making the parasitic-signal-de-embedded s-parameters highly distorted. The best-optimized model s-parameters then still poorly fit the distorted s-parameters, for our model does not take into account the distortion effect (see Sec. 4.1.3.4 for an in-depth discussion). The discrepancy in the hole-doped region is similarly explained, as the measured signal is distorted in a way that cannot be fully captured by the model in use. This distortion can be traced back to the asymmetric behavior caused by work function mismatch in our edge contacts, where the contact between the metal and hole-doped graphene has been demonstrated to exhibit non-ideal behaviors [81] that are difficult to capture with a passive linear model (see Sec. 4.1.3.4).

Back in the higher-fidelity electron-doped region (unshaded region in Fig. 4.5), the data at 296 K result in more appreciable deviation from theory, due to the $\sim 4\times$ decrease in mobility ($\sim 4\times$ increase in R), which reduces all of ϕ_2/ϕ_1, ϕ_2, $\sqrt{C/R}$, and transmission amplitude. This highlights the challenge in measurements of sub-unit Q devices. Nonetheless, while not as quantitatively accurate as the 30-K data, the 296-K data still present a firm direct proof of $L_{\rm k}$ and collective dynamical mass, made possible by the h-BN graphene interface and the proximate gating. Thus even the 296-K data represent a significant leap from the prior works that have only failed to observe the kinetic inductance [78–80].

Beside its fundamental importance for graphene electrodynamics and plasmonics, this work may offer exciting technological vistas. The graphene kinetic inductance as a manifestation of the collective inertia effect is orders of magnitude larger than the magnetic inductance at similar dimensions (see

Sec. 4.1.2.3), and thus can be used in the future to substantially miniaturize inductors, as it allows one to obtain the same inductance value in orders of magnitude smaller area. Radio-frequency integrated circuits, such as resonators, filters, oscillators, and amplifiers, prevalent in communication and computing systems, suffer from large chip areas due to magnetic inductors. Thus these high frequency applications may benefit greatly from harnessing the kinetic inductance of graphene revealed in this work. Better room temperature scalability and facile tunability as compared to traditional kinetic inductors from superconductors and semiconductor two-dimensional gases also bode well in this direction as the mobility of graphene continues to improve. Furthermore, the bias-dependency of graphene kinetic inductance renders graphene a natural voltage-controlled tunable inductor as a counterpart to the prevalent voltage-controlled semiconductor capacitor.

4.1.2.3 Comparison of Kinetic and Magnetic Inductance

We envisioned that the large kinetic inductance of graphene compared to its magnetic inductance may enable one to build substantially miniaturized inductors in orders of magnitude smaller area. This can be more concretely seen in numbers in the following examples.

Magnetic inductors prevalent in high frequency analog integrated circuits occupy very large chip area. For instance, in order to obtain 2.4 nH of inductance at GHz frequencies, around $4\ \mathrm{mm}^2$ of chip area is needed [87]. Taking the kinetic inductance per square of graphene as $100\ \mathrm{pH/square}$ (at ~ 5 V bias in Fig. 4.5a), with a graphene width of $10\ \mu\mathrm{m}$, the same amount of inductance would be obtained in a $0.0024\ \mathrm{mm}^2$ area, which is about 2600 times smaller than the magnetic inductor considered earlier. Such an application may be a possibility as the mobility of CVD-grown graphene continues to improve.

As another example of the comparison between the magnitudes of kinetic and magnetic inductances in graphene, we can consider our own graphene device of Fig. 4.3. The magnetic inductance per unit length in a parallel plate configuration (graphene and the top gate) is given by $L_\mathrm{m} = \mu\frac{d}{W}$, where μ is the magnetic permeability, d is the distance between the plates, and W is the width of the plates. With $d \approx 200$ nm and $W = 7.5$ um in our device, this magnetic inductance per unit length evaluates to $3 \times 10^{-3}\ \mathrm{pH}/\mu\mathrm{m}$, while the kinetic inductance per unit length (kinetic inductance per square divided by the width, W) is measured to be around $10\ \mathrm{pH}/\mu\mathrm{m}$, 3 orders of magnitude larger than the magnetic inductance. The magnetic inductance was ignored in the model of Fig. 4.1d for this reason.

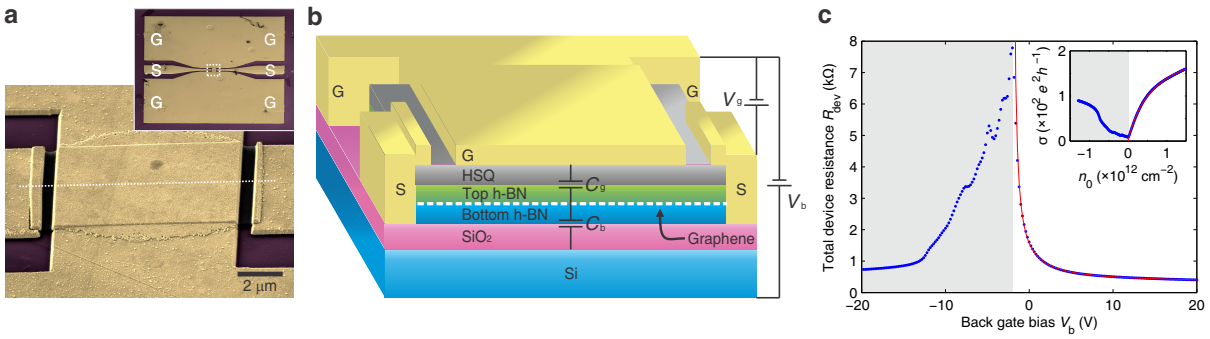

Figure 4.6: Device description and DC measurements of additional device [16]. **a**, False colored scanning electron micrograph of the device. The rough, rounded edges around the central rectangular region are artifacts from fabrication not important in the analysis. **b**, Schematic illustration of the device. Device dimensions are $W = 3.5\ \mu$m, $l = 8.9\ \mu$m, top h-BN thickness 43 nm, bottom h-BN thickness 13 nm, and HSQ thickness ~ 100 nm. **c**, DC total device resistance $R_{\rm dev}$ measured at 30 K as a function of $V_{\rm b}$, while graphene and the top gate are kept at the same DC potential, *i.e.*, $V_{\rm g} = 0$ (inset: corresponding plot of conductivity estimate, $\sigma \equiv (R_{\rm dev}/(l/W))^{-1}$; $n_0 = C_{\rm b}/W \times (V_{\rm b} - V_{\rm b,0})/e$ with $C_{\rm b}/W = 0.12$ fF/μm^2 and $V_{\rm b,0} = -2\ V$). Red solid curves are fits to $\sigma^{-1} = (n_0 e \mu_{\rm C})^{-1} + \rho_{\rm s}$ with $\mu_{\rm C} = 85,000$ cm^2/Vs and $\rho_{\rm s} = 113\ \Omega$.

4.1.2.4 Additional Device Measurements

Here we present measurement results from another device in addition to the device of Fig. 4.3. The results confirm that the measurement and analysis presented in this work are clearly reproducible despite the difficult microwave measurement conditions detailed in Sec. 4.1.1.1, owing to the h-BN encapsulation of graphene, one-dimensional edge contact, low temperature, and delicate microwave phase measurements.

Fig. 4.6 shows the additional device's image and its DC 2-terminal resistance measurement result at 30 K. The device is relatively smaller ($W = 3.5\ \mu$m, $l = 8.9\ \mu$m) compared to the device of Fig. 4.3, and shows a lower mobility of $\mu_{\rm C} = 85,000$ cm^2/Vs in the electron-doped region ($V_{\rm b} > V_{\rm b,0} = -2$ V). The hole-doped region ($V_{\rm b} < V_{\rm b,0}$) shows a much stronger asymmetric behavior compared to the device of Fig. 4.3, and therefore we focus our analysis on the electron-doped region where it functions as a clean graphene device.

Fig. 4.7 shows the results from microwave measurements performed with this device at 30 K. For this device, bias voltage is applied on the S-lines of the CPWs via bias tees ($V_{\rm g}$), while the back gate was kept at the same DC potential as the top gate ($V_{\rm b} = 0$). This means the charge density induced on graphene is now expressed as $n_0 \approx (C_{\rm b} + C_{\rm g})/W \times (V_{\rm g,0} - V_{\rm g})/e$, as opposed to $n_0 \approx C_{\rm b}/W \times (V_{\rm b} - V_{\rm b,0})/e$ for the DC 2-terminal measurement. $C_{\rm g}/W$ is estimated to be roughly 0.22 fF/μm^2 from the thicknesses of the top h-BN and the HSQ layer. The microwave measurement results seen

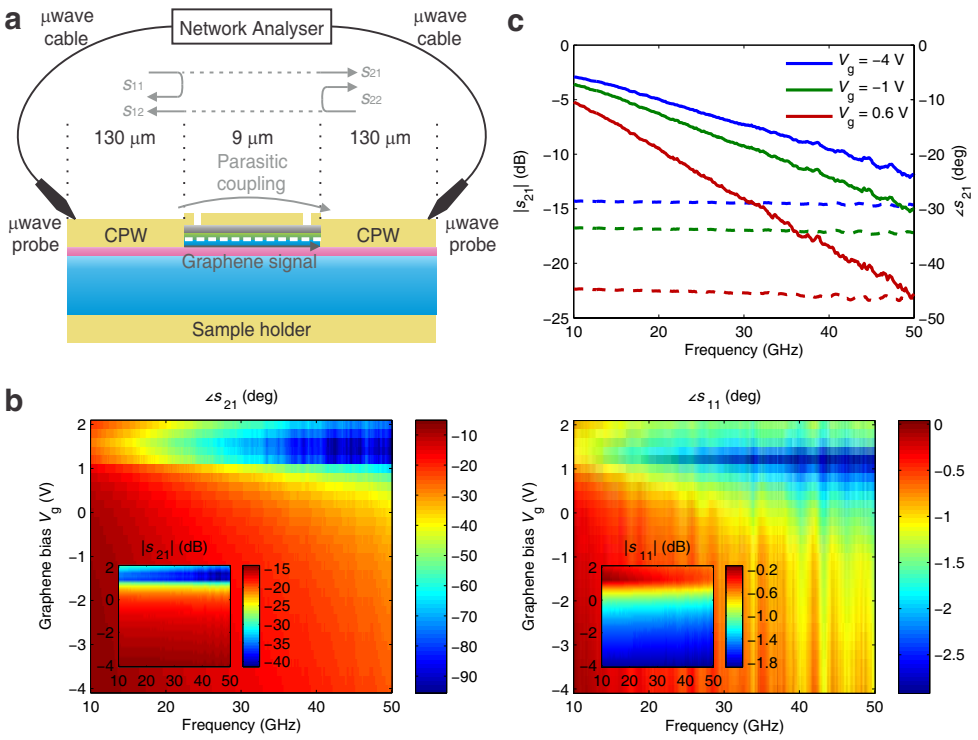

Figure 4.7: Microwave s-parameter measurements of additional device [16]. **a**, Schematic diagram of the measurement setup. The s-parameters shown are after calibrating out the delay and loss of the cables, probes, and on-chip CPWs, and also after de-embedding the parasitic coupling bypassing graphene. **b**, Phase (insets: amplitude) of the measured transmission (s_{21}; left) and reflection (s_{11}; right) parameters after the calibration and de-embedding at 30 K. For this device, bias voltage is applied on the S-lines of the CPWs via bias tees ($V_{\rm g}$), while the back gate was kept at the same DC potential as the top gate ($V_{\rm b} = 0$). **c**, Select data from **b**, specifically, transmission phase ($\angle s_{21}$; solid curves) and amplitude ($|s_{21}|$; dashed curves) at three representative bias values $V_{\rm g}$ = -4, -1, and 0.6 V ($V_{\rm b} = 0$ V).

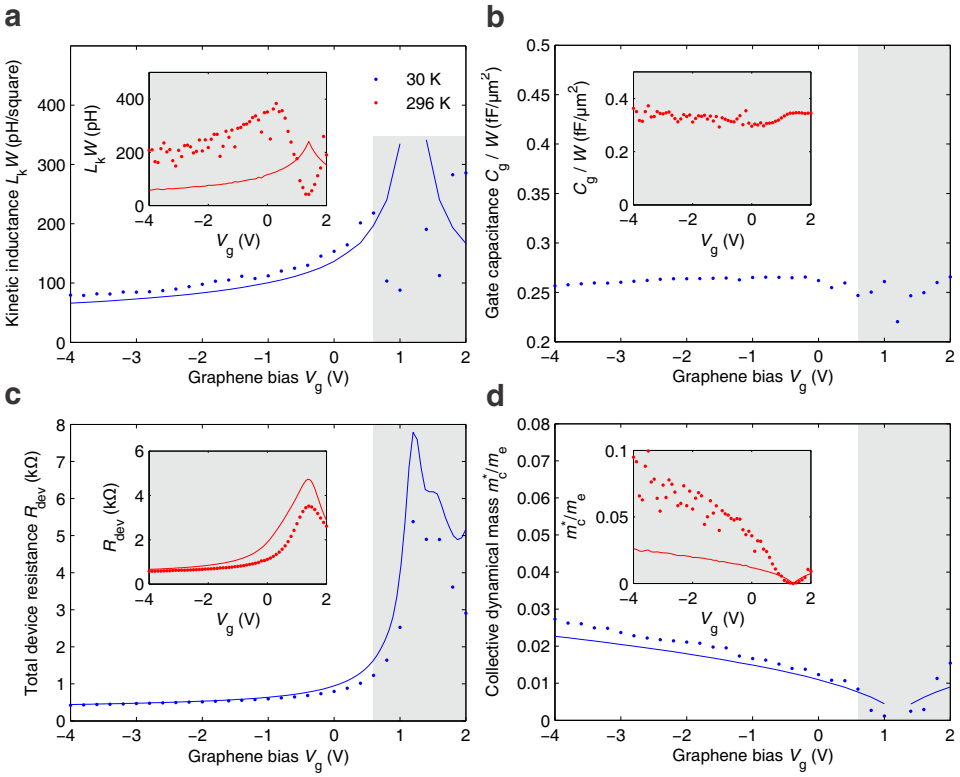

Figure 4.8: Kinetic inductance per square, $L_{\mathrm{k}}W$ (**a**), graphene to top-gate capacitance per area, C_{g}/W (**b**), total device resistance, R_{dev} (**c**), and collective dynamical mass per electron, m_{c}^* (**d**), extracted from the measured s-parameters for various V_{g} at 30 K and 296 K [16]. The solid curves in **a** and **d** represent theoretical predictions. The solid curve in **c** is R_{dev} measured at DC (Fig. 4.6c) but with the x-axis inverted and rescaled according to the ratio of the capacitance C_{b} relevant to the DC measurement of Fig. 4.6c, to the capacitance $C_{\mathrm{g}} + C_{\mathrm{b}}$ relevant to the DC biasing in the microwave measurements.

in Figs. 4.7b,c are qualitatively similar to those obtained from the device of Fig. 4.3, amenable to L_{k} extraction.

Fig. 4.8 shows the device model parameters (L_{k}, C_{g}, and R_{dev}) extracted from the microwave measurement data. The results are qualitatively and quantitatively very similar to the device of Fig. 4.3. In the electron-doped region away from charge neutrality ($V_{\mathrm{g}} < V_{\mathrm{g},0} = 1.2$ V; note the inverted direction due to the different biasing scheme in this measurement), extracted C_{g} stays nearly constant close to the expected value, extracted L_{k} closely follows the theoretically expected curve, and R_{dev} extracted from microwave measurements matches that measured at DC. The collective mass m_{c}^* obtained from L_{k} also closely follows the theoretically expected curve. This confirms that the measurement and analysis presented in this work are clearly reproducible despite the difficult microwave measurement conditions posed by in-graphene and contact resistances. The analysis in the following (Sec. 4.1.3) will be based on

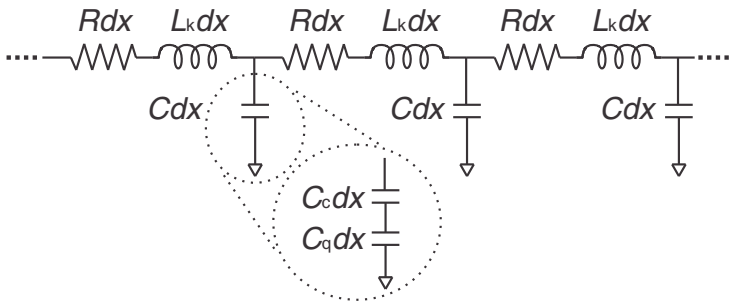

Figure 4.9: Lossy transmission line model for proximately gated graphene [16].

this device unless noted otherwise.

4.1.3 s-Parameters and Extraction of L_k

4.1.3.1 Phase Delay $\phi(L_\mathrm{k}, C, R)$ and Device Design Implications

In Sec. 4.1.2, the propagation phase delay ϕ expressed in terms of L_k, C, and R provided the key device design guideline (encapsulation of graphene with top and bottom h-BN layers, and proximate top gating) to enable the L_k measurement. Here we derive this ϕ expression, and elaborate more on the design guideline. As the electron scattering severely interferes with L_k measurement, we consider the full graphene transmission line model of Fig. 4.1d, including per-unit-length resistance R modeling the electron scattering effect. For convenience, this lossy transmission line is re-produced in Fig. 4.9.

Let a wave of a microwave angular frequency ω propagating on the graphene transmission line be represented by the phasor $e^{-\gamma z}$ with the complex propagation factor $\gamma = \alpha + i\beta$ (α, β are real). γ's real part, α, captures the loss in the transmission line. Its imaginary part, β, is actually the plasmonic wavenumber k_p, as the wave in the graphene transmission line model considered here is the graphene plasmonic wave. As was seen in Sec. 2.3.1, γ is related to the transmission line's per-unit-length components, L_k, C, and R, as

$$\gamma = \alpha + i\beta = \sqrt{(R + i\omega L_\mathrm{k})(i\omega C)}. \tag{4.17}$$

The kinetic inductor's quality factor $Q = \omega L_\mathrm{k}/R$ is smaller than 1. Even after we substantially reduce R with the h-BN encapsulated structure and 30-K operation, Q ranges from 0.05 to 0.2 for the device of Fig. 4.6 and from 0.2 to 0.8 for the device of Fig. 4.3, as frequency is varied from 10 to 50 GHz; with graphene on a more standard substrate such as SiO_2, R is far larger and Q is even smaller. Therefore,

we can approximate[4] the expression above to the first order of $Q = \omega L_{\mathrm{k}}/R$ as

$$\begin{aligned}\alpha &\approx \sqrt{\frac{\omega RC}{2}}\left(1 - \frac{1}{2}\frac{\omega L_{\mathrm{k}}}{R}\right)\\ \beta &\approx \sqrt{\frac{\omega RC}{2}}\left(1 + \frac{1}{2}\frac{\omega L_{\mathrm{k}}}{R}\right).\end{aligned} \tag{4.18}$$

The total propagation phase delay through the graphene transmission line of length l is βl, and thus, the per-unit-length phase delay ϕ is no more than β, and we express it as in Sec. 4.1.2.2:

$$\phi \approx \underbrace{\sqrt{\frac{\omega RC}{2}}}_{\phi_1} + \underbrace{\sqrt{\frac{\omega^3}{8}}\sqrt{\frac{C}{R}}L_{\mathrm{k}}}_{\phi_2}. \tag{4.19}$$

As seen, while the first term ϕ_1 is independent of L_{k}, the second term ϕ_2 contains L_{k}, and thus, is of key interest; incidentally, their ratio is given by

$$\frac{\phi_2}{\phi_1} = \frac{1}{2}\frac{\omega L_{\mathrm{k}}}{R} = \frac{Q}{2}. \tag{4.20}$$

As described in Sec. 4.1.2.2, decreasing R and increasing C are crucial for a given L_{k} to have a more 'measurable' impact on the phase delay ϕ, whose information is essentially included in the transmission coefficients s_{21} and s_{12}, which will be discussed in more detail in Sec. 4.1.3.2. The R-reduction proportionally improves $\phi_2/\phi_1 = Q/2$ and makes ϕ_2 a more appreciable fraction of ϕ_1, by reducing ϕ_1 and amplifying ϕ_2. The C-enhancement keeps ϕ_2/ϕ_1 constant, but still increases ϕ_2 itself. Taken together, the R-reduction and C-enhancement amplify ϕ_2 beyond the phase measurement error, which we call ϕ_{e}, caused by the imperfect calibration and non-ideal parasitic signal de-embedding. Now, the criterion $\phi_2 > \phi_{\mathrm{e}}$ we have focused right above is necessary but not sufficient for L_{k} extraction. $\Delta\phi_2 > \phi_{\mathrm{e}}$ must be also satisfied, where $\Delta\phi_2$ is the variation of ϕ_2 corresponding to a target L_{k} extraction accuracy (resolution) ΔL_{k}, *i.e.*, $\Delta\phi_2 = \sqrt{\omega^3/8}\sqrt{C/R}\Delta L_{\mathrm{k}}$. To meet this additional criterion, we have to maximize $\Delta\phi_2/\Delta L_{\mathrm{k}} = \phi_2/L_{\mathrm{k}} \propto \sqrt{C/R}$, which is also achieved by the R-reduction and C-increase; in fact, the R-reduction and C-enhancement increased ϕ_2 above, by increasing the proportionality factor

[4]The approximation, which may be numerically inaccurate near 50 GHz for the device of Fig. 4.3 due to its large mobility, is used here to capture the most dominant effect affecting the measurements without complicating the algebra. However, no approximation is used in the extraction procedure (Sec. 4.1.3.3) to ensure accuracy.

$\sqrt{C/R}$.

To substantially reduce R, we interface graphene with h-BN layers on both the top and bottom sides, and to obtain extra R-reduction, we also lower the operation temperature to 30 K in our experiment. The C-enhancement is achieved by the proximate top gating. As discussed in Sec. 2.1.2.1, with the distance d between graphene and top gate being much smaller than graphene plasmonic wavelength $\lambda_{\rm p} = 2\pi/k_{\rm p}$ (*i.e.*, $k_{\rm p}d \ll 2\pi$, which is the case with our device), $C_{\rm c}$ of Fig. 4.9 is just the parallel plate capacitance, $C_{\rm c} = \kappa\epsilon_0 W/d$, and with the effect of $C_{\rm q}$ negligible, $C = C_{\rm c} = \kappa\epsilon_0 W/d$. This is much larger than the capacitance of ungated graphene[5] $2\kappa\epsilon_0 k_{\rm p} W$ because $k_{\rm p}d \ll 2\pi$. We can indefinitely increase C of our gated structure by keeping reducing d, but we stop at a certain point; in fact, we placed the extra layer of HSQ in addition to the top h-BN layer between graphene and top gate so that d is not too small. This is because with too large a C value, the attenuation constant $\alpha \propto \sqrt{C}$ of Eq. (4.18) would become too excessive, causing a significant attenuation. The C value chosen in our work is large enough to enable $L_{\rm k}$ extraction, but not so large so that we can maintain mild attenuation; $\alpha l \approx l\sqrt{\omega RC/2}$ ranges around $0.1 \sim 2$, depending on frequency ω and graphene bias $V_{\rm g}$, as far as we keep away from the neutrality point, *e.g.*, up to $V_{\rm g} \sim 0.6$ V for the device of Fig. 4.6.

4.1.3.2 Detailed Analysis of s_{21}-Parameters

To confirm that the C-enhancement and R-reduction indeed make the measured s-parameters amenable to $L_{\rm k}$ extraction, we analyze in details the measured transmission (s_{21}) parameters in Fig. 4.7c in conjunction with simulations. In the foregoing section, we discussed the impact of the C-enhancement and R-reduction not on s_{21}'s phase ($\angle s_{21}$), but on the propagation phase delay ϕl. These two phase quantities are not exactly the same, because $\angle s_{21}$ takes into account not only ϕl, but also the phase change incurred by the reflection at the CPW-graphene interface[6]. Nonetheless, the behavior of ϕl is strongly reflected in $\angle s_{21}$, and thus, the impact of C-enhancement and R-reduction on ϕl should be also distinctively observed from $\angle s_{21}$. With this understanding, in the analysis of the s_{21} parameters here, our language will

[5] In either our top gated case or the ungated case imagined here with our device, the back gate *un*connected to the G lines of the CPWs in our device is irrelevant as far as the microwave signaling is concerned, thus the capacitance $C_{\rm b}$ associated with the back gate does not come into our consideration here.

[6] More concretely, s_{21} can be approximated as the following, after ignoring multiple reflection effects and contact effects for simplicity [59]:

$$s_{21} \approx \frac{4R_0 Z_0}{(R_0 + Z_0)^2} e^{-\alpha l} e^{-i\beta l}. \tag{4.21}$$

not be too rigorous in distinguishing the two phase quantities; we seek to present the essence instead of the most rigorous analysis that complicates algebra.

To appreciate the impact of the C-increase and R-reduction on our ability to extract $L_{\rm k}$, we compare the measured s_{21}-parameters to the s_{21}-parameters simulated under various scenarios. For the s-parameter simulation, we use Sonnet frequency-domain electromagnetic field solver, where the graphene is modeled as a two-dimensional conductor where its resistive and kinetic inductive impedances enter as simulation parameters. Its capacitance (and negligible magnetic inductance) is attained as part of the simulation outcome. Electromagnetic waves in the frequency range of 10-50 GHz are launched onto the CPWs in the simulator; the simulated response of the graphene device is recorded in terms of s-parameters at each frequency.

1. **Reconstruction of the measured s_{21} of Fig. 4.7c:** Fig. 4.10a shows $\angle s_{21}$ (solid curves) and $|s_{21}|$ (dashed curves) simulated for our top-gated, h-BN encapsulated graphene device. This simulation is done with three sets of $L_{\rm k}$ and R values, extracted[7] from the measured s-parameters—whose s_{21} is in Fig. 4.7c—at three different graphene bias ($V_{\rm g}$) values. The simulated s_{21} of Fig. 4.10a is almost identical to the measured s_{21} of Fig. 4.7c. We can also examine the current distribution in the CPWs and graphene; a simulated graphene-layer current distribution example corresponding to the red curves of Fig. 4.10a at 50 GHz is presented in Fig. 4.10b, visualizing the signal propagation from the left CPW through graphene to the right CPW with attenuation.

2. **$L_{\rm k}$'s contribution to s_{21}:** The dark-colored simulated s_{21} curves of Fig. 4.10c are the repetition of Fig. 4.10a, but the light-colored s_{21} curves of Fig. 4.10c are simulated after removing $L_{\rm k}$ from the impedance parameter of graphene used in the simulation. The appreciable change in $\angle s_{21}$ curves after removing $L_{\rm k}$ at each bias reflects that the $L_{\rm k}$-bearing ϕ_2 term is a measurable fraction of the $L_{\rm k}$-independent ϕ_1 term. This is owing to the reduced R and increased C in our device.

Here $R_0 = 50\ \Omega$ is the characteristic impedance of the measurement environment, and Z_0 is the characteristic impedance of the lossy graphene transmission line,

$$Z_0 = \sqrt{\frac{R + i\omega L_{\rm k}}{i\omega C}} \approx \sqrt{\frac{R}{2\omega C}} \left[\left(1 + \frac{1}{2}\frac{\omega L_{\rm k}}{R} \right) - i \left(1 - \frac{1}{2}\frac{\omega L_{\rm k}}{R} \right) \right]. \tag{4.22}$$

where the last expression is an approximation to the first order of $Q = \omega L_{\rm k}/R$. As can be seen, $\angle s_{21}$ is not just $\phi l = \beta l$ but includes the phase change associated with the reflection, captured by the complex factor $4R_0 Z_0/(R_0 + Z_0)^2$.

[7] The extraction procedure will be fully detailed in Sec. 4.1.3.3.

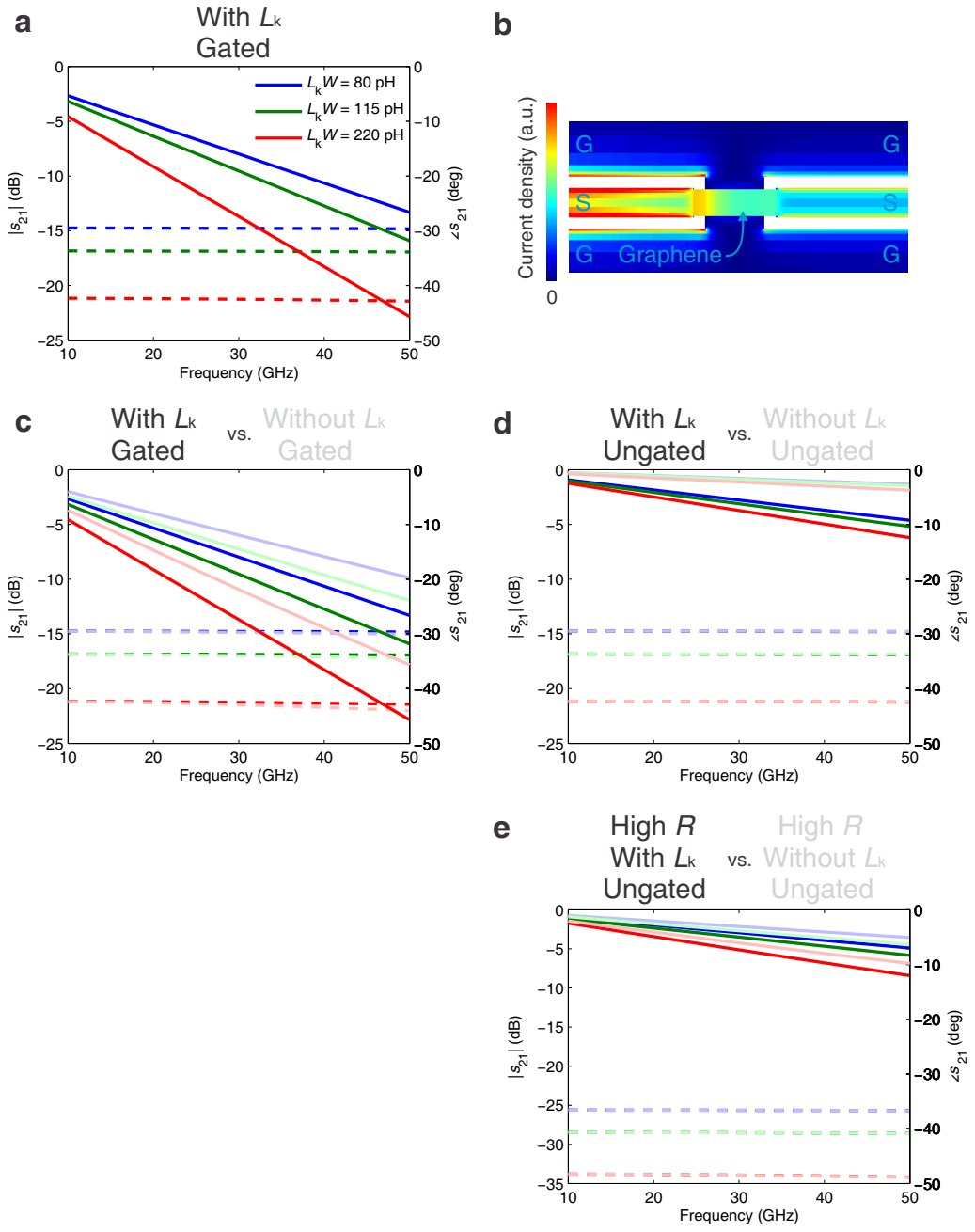

Figure 4.10: Analysis of simulated s_{21}-parameters [16]. **a**, Simulated $\angle s_{21}$ (solid curves) and $|s_{21}|$ (dashed curves) for the gated h-BN interfaced graphene device in Fig. 4.7 with $(L_kW,\ RW) = (80\ \text{pH},\ 140\ \Omega)$ [blue], $(115\ \text{pH},\ 200\ \Omega)$ [green], and $(220\ \text{pH},\ 370\ \Omega)$ [red] per square, and contact resistances of 53 Ω on each side. **b**, Simulated current density distribution in the graphene layer at 50 GHz in the red-colored case of **a**. **c**, Dark-colored curves are identical to **a**; light-colored curves are simulations without L_k in otherwise the same situation as **a**. **d**, Simulation results after removing the top gate from the case of **c**. **e**, Simulation results after increasing R by 5 times at each bias from the case of **d** (*i.e.*, RW is 700 Ω [blue], $1,000\ \Omega$ [green], and $1,850\ \Omega$ [red] per square).

3. **Impact of C-increase in our device:** The s_{21} curves in Fig. 4.10d are simulated after removing the proximate top gate (thus with decreased C) in otherwise the identical simulation settings as Fig. 4.10c. As seen in Fig. 4.10d, even with the lower C, $\angle s_{21}$ curves before and after removing $L_{\rm k}$ at any given bias exhibit an appreciable difference, because C does not affect ϕ_2/ϕ_1 (Eq. (4.20)). On the other hand, with C reduction, the progression of $\angle s_{21}$ with frequency (and thus $\angle s_{21}$ itself) substantially decreases, as can be seen by comparing dark $\angle s_{21}$ curves between Fig. 4.10c and Fig. 4.10d. Consequently, without the top gate, the variation of $\angle s_{21}$ with varying values of $L_{\rm k}$ due to different graphene biases reduce to $\sim 1^\circ$ even at the highest frequency (dark $\angle s_{21}$ curves of Fig. 4.10d), while the phase measurement accuracy $\phi_{\rm e}$ in our microwave measurement is typically limited to $\sim 1^\circ$ at best, due to the (inherently) imperfect calibration and non-ideal parasitic signal de-embedding[8]. This shows how the top gating and consequently larger C in our device enables $L_{\rm k}$ extraction.

4. **Impact of R-reduction in our device:** The s_{21} curves of Fig. 4.10e are simulated without top gating, just as in the case of Fig. 4.10d, and now also with 5 times larger R value at each bias to emulate the situation of graphene interfaced with a more standard substrate (*e.g.*, SiO_2/Si) and thus with reduced mobility. The already bad situation of Fig. 4.10d is now even worsened in Fig. 4.10e, where the dark-colored $\angle s_{21}$ curves with $L_{\rm k}$ and light-colored $\angle s_{21}$ curves without $L_{\rm k}$ at each bias become close with difference $\sim 1^\circ$ even at the highest frequency. This simulation demonstrates how the smaller R in our device helps $L_{\rm k}$ extraction.

5. **Behavior of $|s_{21}|$:** So far we have focused on $\angle s_{21}$, but $|s_{21}|$ is also of importance. As can be seen in and across Figs. 4.10c,d, $|s_{21}|$ is hardly affected by $L_{\rm k}$ or C, but is almost solely determined by R. Specifically: when $L_{\rm k}$ is removed, $|s_{21}|$ at a given bias remains almost the same in either Fig. 4.10c or Fig. 4.10d; with differing C values between Figs. 4.10c and d, $|s_{21}|$ at a given bias also remains practically the same; by contrast, both Figs. 4.10c and d show that with increasing R with the varying graphene bias, $|s_{21}|$ conspicuously decreases. This R dependency of $|s_{21}|$ can be also seen by comparing Figs. 4.10c,d with Fig. 4.10e; with the 5 times larger R at any given bias, $|s_{21}|$ in Fig. 4.10e is conspicuously smaller than $|s_{21}|$ in Figs. 4.10c,d. Too small a value of

[8] Section 4.1.3.5 will present our experiment with an ungated graphene device, demonstrating the exceeding difficulty in $L_{\rm k}$ extraction from the s-parameters in the ungated case.

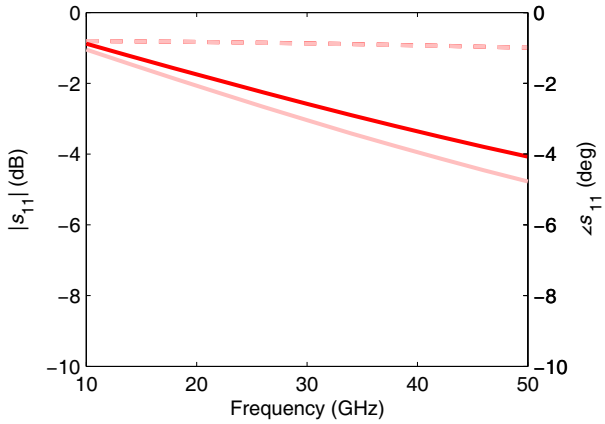

Figure 4.11: Simulated $\angle s_{11}$ (solid curves) and $|s_{21}|$ (dashed curves) with and without L_k (dark and light-colored), which corresponds to the dark and light red-colored simulated s_{21} curves of Fig. 4.10c [16].

$|s_{21}|$ as in Fig. 4.10e (or near the charge neutrality point not discussed in this section) makes the de-embedding of graphene-bypassing parasitic signal highly error-prone, leading to spurious L_k, as will be discussed in Sec. 4.1.3.3. This is another reason why we should reduce R, hence the necessity of our h-BN graphene interface.

In the above, we have shown how reduced R and increased C allow L_k to exert a more measurable impact on s_{21}. However, this does not mean that L_k can be extracted solely from s_{21} (and s_{12})[9]. While s_{21} certainly carries the information on L_k, L_k cannot be determined separately from C with s_{21} alone, because the effects of L_k and C are mixed in $\angle s_{21}$, and they have little impact on $|s_{21}|$. To determine L_k and C separately, we also need the reflection parameter s_{11} (and s_{22}).

We can see the effects of L_k and C on s_{11} from the Sonnet electromagnetic simulation of our top-gated, h-BN encapsulated graphene device; Fig. 4.11 shows the simulated s_{11} with and without L_k (dark and light-colored, respectively), which correspond to the dark and light red-colored simulated s_{21} curves of Fig. 4.10c. By comparison, we can see that while L_k and C had an additive effect on $\angle s_{21}$ (they both increased $\angle s_{21}$), they have a subtractive effect for $\angle s_{11}$ (C increases $\angle s_{11}$ but L_k decreases $\angle s_{11}$). Therefore, by combining $\angle s_{21}$ and $\angle s_{11}$ measurements, L_k and C can be separately determined.

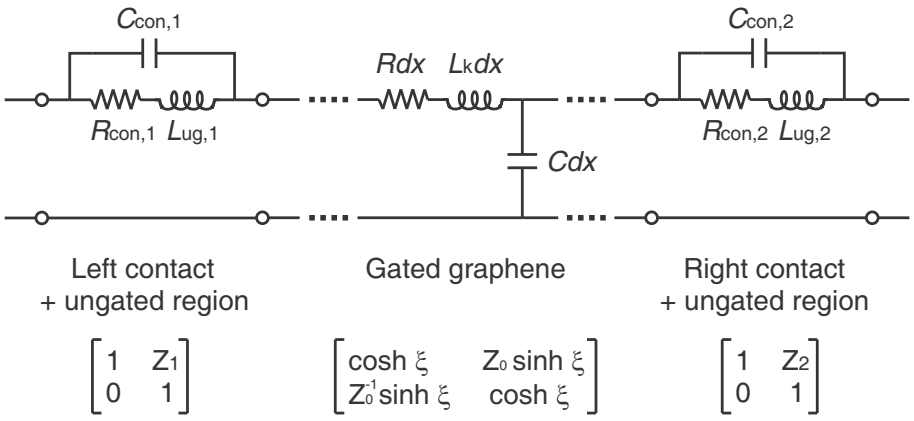

Figure 4.12: Model used for fitting to the measured s-parameters [16]. Corresponding transmission matrix representation is shown below. Here, $\xi = l\sqrt{(R + i\omega L_{\rm k})(i\omega C)}$ [dimensionless], $Z_0 = \sqrt{(R + i\omega L_{\rm k})/(i\omega C)}$ [Ω], and $Z_{1,2} = (R_{\rm con,1,2} + i\omega L_{\rm ug,1,2})||(1/i\omega C_{\rm con,1,2})$ [Ω]. R, $L_{\rm k}$, and C are per-unit-length variables whereas $R_{\rm con,1,2}$, $L_{\rm ug,1,2}$, and $C_{\rm con,1,2}$ are lumped variables.

4.1.3.3 Device Parameter Extraction Procedure

Figure 4.12 shows the graphene transmission line model, plus the models for the left and right graphene contacts with the CPWs' S-lines. The left [right] contact model consists of $R_{\rm con,1}$ [$R_{\rm con,2}$] accounting for the contact resistance as well as the resistance of the small ungated graphene region near the left [right] contact seen in Figs. 4.3 or 4.6, $L_{\rm ug,1}$ [$L_{\rm ug,2}$] accounting for the kinetic inductance of the small ungated graphene region on the left [right], and capacitance $C_{\rm con,1}$ [$C_{\rm con,2}$] due to the small segment of the left [right] S-line edging over graphene. For a given set of model parameters ($L_{\rm k}$, C, R and the contact model component parameters), we calculate the s-parameters using the transmission matrix method [25]; this is a precise calculation, contrasting the approximate calculations that appeared in Secs. 4.1.3.1 and 4.1.3.2 to illustrate the physics of the measurement in a simple manner. The calculated model s-parameters consist of 8 sets of curves (real and imaginary parts of s_{11}, s_{21}, s_{12}, and s_{22}) that span the frequency range of 10-50 GHz. To determine $L_{\rm k}$ as well as other model parameters at a given bias, we repeat the calculation by altering the model component parameters until the calculated model s-parameters best fit the measured s-parameters in the sense of least square curve fit, by using 'lsqcurvefit' function of MATLAB. This procedure is repeated at each bias voltage.

This optimization procedure requires a set of initial guesses for each model parameter. To ensure no arbitrariness, the same set of initial guesses were used across all the different sets of measurement data

[9] If our device is perfectly reciprocal, $s_{21} = s_{12}$; in reality, the perfect reciprocity is somewhat compromised, because the left and right contacts can behave differently.

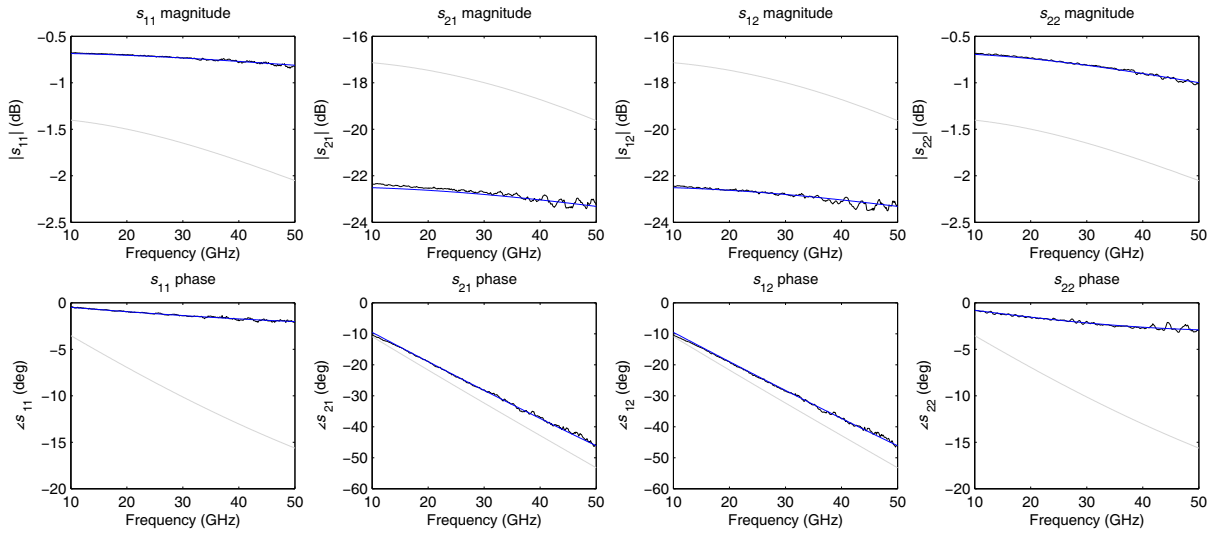

Figure 4.13: Measured (black) *vs.* fitted (blue) s-parameters for the device data of Fig. 4.8 at $V_{\rm g} = 0.6$ V (30 K) [16]. Gray s-parameter curves obtained with the initially-guessed model parameters evolve to the blue curves as the optimization proceeds.

taken at varying bias voltages and temperatures for both device of Figs. 4.3 and 4.6. These initial guesses are also very generic, taking values such as $L_{\rm k}W = 50$ pH, $Rl = 500\ \Omega$, and $Cl = 1$ fF. The fitting results are insensitive to the initial guesses; for instance, providing initial values 10$\sim$20 times away from the actual values does not alter the end result. The maximum and minimum bounds of the optimization range for each model parameter, also needed by the 'lsqcurvefit' function, were set well away from the parameter's expected end value to ensure no interference with the arbitrarily set boundaries (*e.g.* $5\ \Omega \leq Rl \leq 50,000\ \Omega$, 0.5 pH $\leq L_{\rm k}W \leq 5000$ pH, 0.01 fF $\leq Cl \leq 100$ fF, etc.). These initial guesses and upper/lower bounds basically serve as a rough estimate of the order of magnitude that the parameters are expected to take for the 'lsqcurvefit' function to facilitate the curve fitting.

As an example, the final curve fits for the device data of Fig. 4.8 at $V_{\rm g} = 0.6$ V (30 K) are shown in Fig. 4.13. The final model s-parameters almost exactly match the measured s-parameters, attesting to the physical validity of the model of Fig. 4.12. The fluctuations in the measured s-parameters in the higher frequency regions are due to residual parasitic signals and calibration errors; as these are not modeled by Fig. 4.12, the model s-parameters do not generate such fluctuations. Device parameters so extracted are the $V_{\rm g} = 0.6$ V (30 K) data points in Fig. 4.8. The rest data points of Fig. 4.8 as well as Fig. 4.5 were obtained through exactly the same procedure.

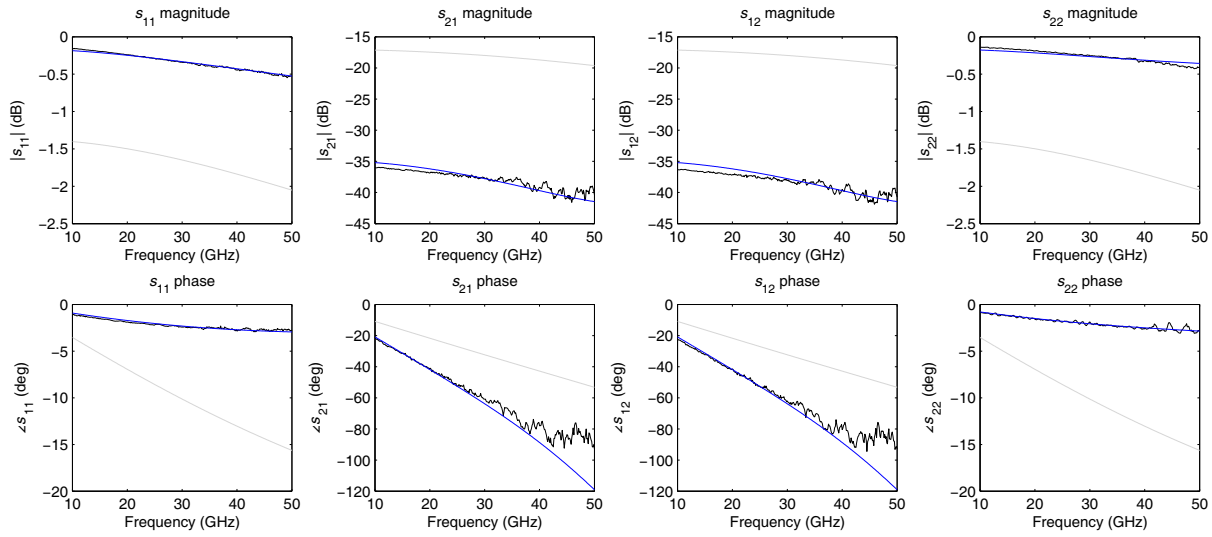

Figure 4.14: Measured (black) *vs.* fitted (blue) s-parameters for the device data of Fig. 4.8 at $V_{\rm g} = 1.2$ V (30 K) [16].

4.1.3.4 Extraction Error Analysis

As seen in Figs. 4.5 or 4.8, the extracted $L_{\rm k}$ value deviates substantially from the theory near the charge neutrality point. In this regime, R is very large, and the transmission through graphene is substantially lowered. Therefore, the *raw* s_{21} and s_{12} before removing the graphene-bypassing parasitic signal are dominated by the parasitic signal itself, rendering the parasitic-signal-de-embedded s_{21} and s_{12} highly distorted with residual parasitic signal. Since our model in Fig. 4.12 does not take into account this residual parasitic signal, the final (best optimized with the least square curve fit) model s-parameters poorly fit the distorted s-parameters. For example, Fig. 4.14 shows the finalized fitting for the s-parameters for the device data of Fig. 4.8 at $V_{\rm g} = 1.2$ V (30 K); the finalized model s_{21} and s_{12} exhibit conspicuous deviation from the measured ones. This explains how the extracted $L_{\rm k}$ at $V_{\rm g} = 1.2$ V (30 K) in Fig. 4.8 becomes spurious, causing its deviation from the prediction.

Even with the bias away from the charge neutrality point, at 296 K, the extracted $L_{\rm k}$ deviates from theory (Fig. 4.5 or inset of Fig. 4.8). In this case, R is increased only by a few times compared to the 30-K case, thus, the detrimental residual parasitic signal effect is not as significant as near charge neutrality point, but the R-increase occurs for a fixed $L_{\rm k}$, reducing both ϕ_2 and ϕ_2/ϕ_1. Consequently, measured s-parameters become once again more fraught with the measurement errors not modeled by Fig. 4.12, rendering $L_{\rm k}$ extraction less accurate. The high sensitivity of our ability to reliably extract $L_{\rm k}$ on R stems

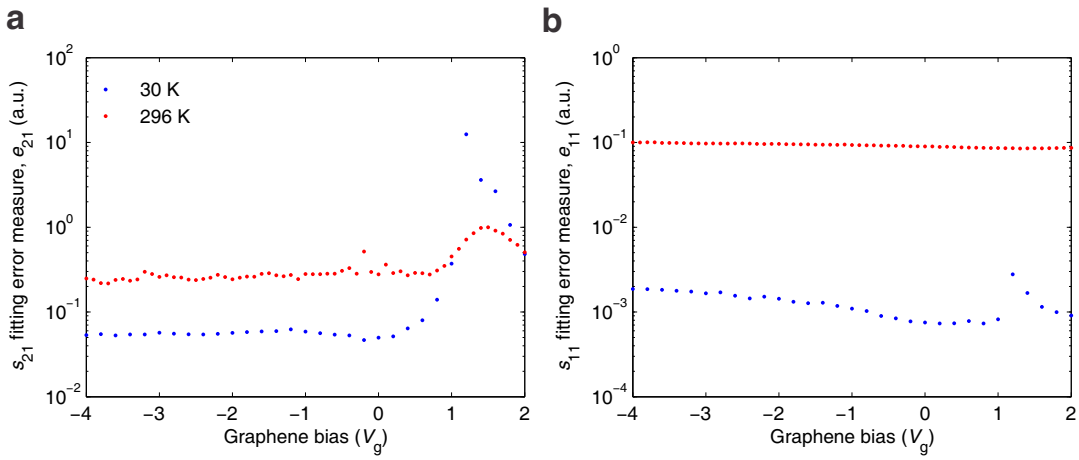

Figure 4.15: Plots of $e_{21}(V_{\rm g})$ and $e_{11}(V_{\rm g})$ for the data corresponding to Fig. 4.8 [16].

from the fact that we are dealing with the sub-unit Q device.

We can quantify the fidelity of the curve fitting for a given type of s-parameter (*e.g.*, s_{21}) by summing the magnitude squared of the residual s-parameter fitting error normalized by the measured s-parameters' magnitude squared, over the measurement frequencies. This quantifies how well the model (Fig. 4.12) is representing the measured data after the optimization. For s_{21}, this measure will be given by

$$e_{21}(V_{\rm g}) = \sum_i \frac{|s_{21,\rm fitted}(f_i, V_{\rm g}) - s_{21,\rm measured}(f_i, V_{\rm g})|^2}{|s_{21,\rm measured}(f_i, V_{\rm g})|^2}, \tag{4.23}$$

where the frequency f_i runs over the measurement frequencies.

Figure 4.15 plots $e_{21}(V_{\rm g})$ and $e_{11}(V_{\rm g})$ for the data corresponding to Fig. 4.8 at 30 K and 296 K. We first note that $e_{11}(V_{\rm g})$ is smaller than $e_{21}(V_{\rm g})$, as in our measurements, $|s_{21}|$ is much smaller than $|s_{11}|$, leaving s_{21} more prone to measurement errors. Next, we note that at 30 K, $e_{21}(V_{\rm g})$ is small away from the charge neutrality point, but becomes very large near the charge neutrality point (note the logarithmic scale); this is consistent with the degree of theory-measurement match of $L_{\rm k}$ shown in Fig. 4.8. The worse theory-measurement match of $L_{\rm k}$ at 296 K as compared to 30 K (Fig. 4.8) in the bias region away from the charge neutrality point is also consistently captured by the fact that $e_{21}(V_{\rm g})$ is larger for the 296 K data in this bias region. All in all, the fully consistent explanation of the sub-optimal-fitting in certain bias and temperature regimes furthers our confidence in the extracted $L_{\rm k}$ values in the regime where the fitting is optimal.

Figure 4.16 plots $e_{21}(V_{\rm b})$ and $e_{11}(V_{\rm b})$ calculated for the data corresponding to Fig. 4.5, showing very

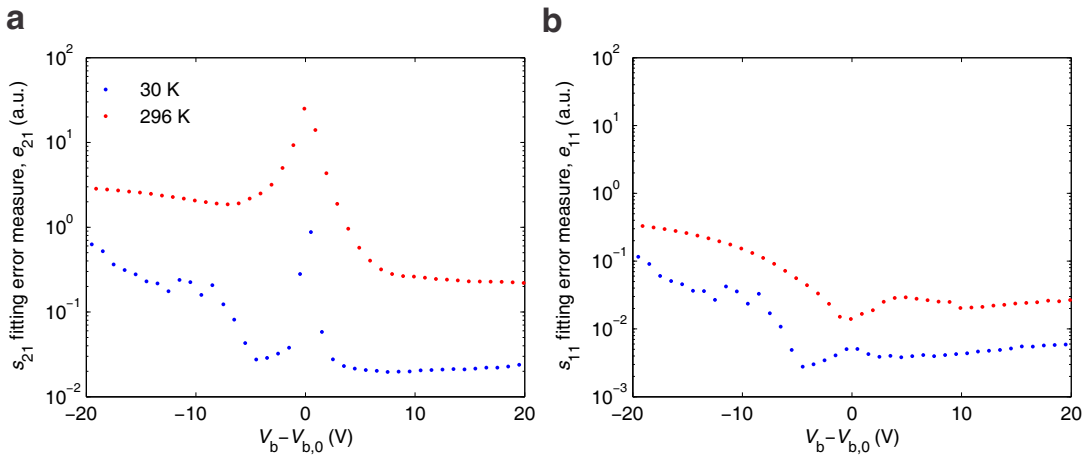

Figure 4.16: Plots of $e_{21}(V_\mathrm{b})$ and $e_{11}(V_\mathrm{b})$ for the data corresponding to Fig. 4.5 [16].

similar behaviors. Specifically: $e_{21}(V_\mathrm{b})$ is the smallest for the electron-doped region ($V_\mathrm{b} > V_{\mathrm{b},0}$) at 30 K where the extracted result was closest to the theory; $e_{21}(V_\mathrm{b})$ shows a spike the near charge neutrality at both temperatures due to the distortion from parasitic signals; $e_{21}(V_\mathrm{b})$ is larger overall for the 296 K data compared to the 30 K data due to the increase in R. Additionally, we can observe that $e_{21}(V_\mathrm{b})$ in the hole-doped region ($V_\mathrm{b} < V_{\mathrm{b},0}$) is conspicuously larger than that in the electron-doped region. This suggests that the microwave measurement data taken in the hole-doped region are distorted in a way that is inexplicable by the model of Fig. 4.12, leading to a larger residual error after the optimization. Clean graphene device fabricated in an identical method [81] is expected to show a nearly symmetric in-graphene resistivity characteristic but the contacts show a highly asymmetric behavior due to the work function mismatch between the metal electrode and hole-doped graphene. Because hole-doped graphene is expected to exhibit exactly the same kinetic inductance and therefore exactly the same model for its microwave characteristics, we suspect that this distortion originates from the contact model of Fig. 4.12 not being an accurate representation of the device with hole-doped graphene. More study will be needed to determine a high-frequency model for the one-dimensional edge contact of hole-doped graphene to metal electrodes that can more accurately describe the measured data.

4.1.3.5 Experiments with Ungated Higher-Resistance Graphene

To further demonstrate how L_k measurement can fail without the strategies we employ (low graphene resistance and top gating), we here present experiments with an ungated graphene device with greater

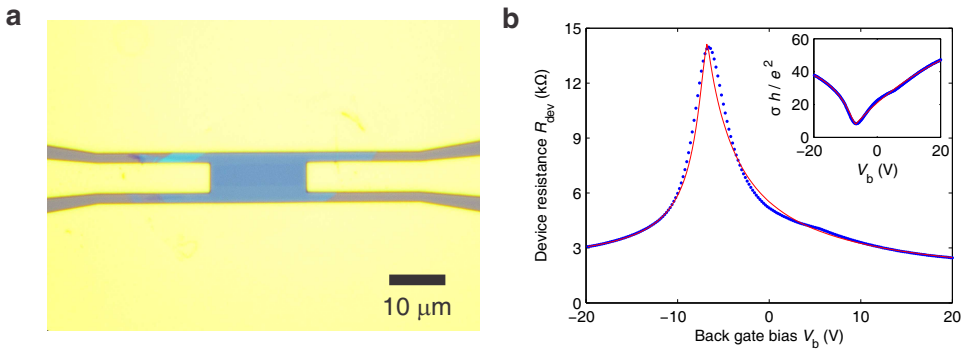

Figure 4.17: DC resistance measurement of ungated graphene device [16]. **a**, Optical image of an ungated graphene device. **b**, 2-terminal DC resistance measurement of the device at 296 K (inset: corresponding conductivity estimate $\sigma \equiv (R_{\rm dev}/(l/W))^{-1}$ normalized to e^2/h).

resistance. The optical image of this device is shown in Fig. 4.17a, where the graphene strip ($W = 4.1\ \mu$m, $l = 18.4\ \mu$m) is visible. 2-terminal DC resistance measurement and analysis similar to those for the device of Fig. 4.3 are performed at 296 K (Fig. 4.17b). Charge neutrality occurs at $V_{\rm b} = V_{\rm b,0} = -6.6$ V, with $\mu_{\rm C} = 8,000\ {\rm cm}^2/{\rm Vs}$ in the electron doped region ($V_{\rm b} > V_{\rm b,0}$) and $\mu_{\rm C} = 26,000\ {\rm cm}^2/{\rm Vs}$ in the hole doped region ($V_{\rm b} < V_{\rm b,0}$). This is considerably lower than the device of Fig. 4.3, due both to the higher temperature and poorer innate sample quality. We also note that the contact resistance of $\sim 2\ {\rm k}\Omega$ in this device is several times worse than the device of Fig. 4.3. The back gate capacitance is almost identical to the device of Fig. 4.3 ($C_{\rm b} = 0.12\ {\rm fF}/\mu{\rm m}^2$).

Microwave s-parameter measurements are performed in the same manner as the other devices. The DC biasing scheme for the microwave measurement is identical to the measurements in Fig. 4.7, but only $C_{\rm b} = 0.12\ {\rm fF}/\mu{\rm m}^2$ is relevant in determining n_0 in this case, as top gate is absent and the aforementioned ungated capacitance $2\kappa\epsilon_0 k_{\rm p} W$ is irrelevant to DC biasing. The results reveal that the device response suffers greatly from parasitic signals (Fig. 4.18) due to the lower mobility and higher contact resistance in this device. Figs. 4.18a,b show measured $|s_{21}|$ and $\angle s_{21}$ after calibration, but before removing the parasitic signals. We see that at certain biases ($V_{\rm g} = 0$ V), the device signal is almost completely buried in parasitic signals, while in other biases the signal is increasingly affected by parasitic signals at high frequencies where the parasitic signal magnitude is larger.

After de-embedding the parasitic signals (Figs. 4.18c,d), a substantial deformation occurs to the measured $\angle s_{21}$ (compare to Fig. 4.4c), especially on the $V_{\rm g} = 0$ V data. Even the less distortion at $V_{\rm g} = 10$ V and 20 V is still quite detrimental. In addition, as graphene is not gated in this device, the

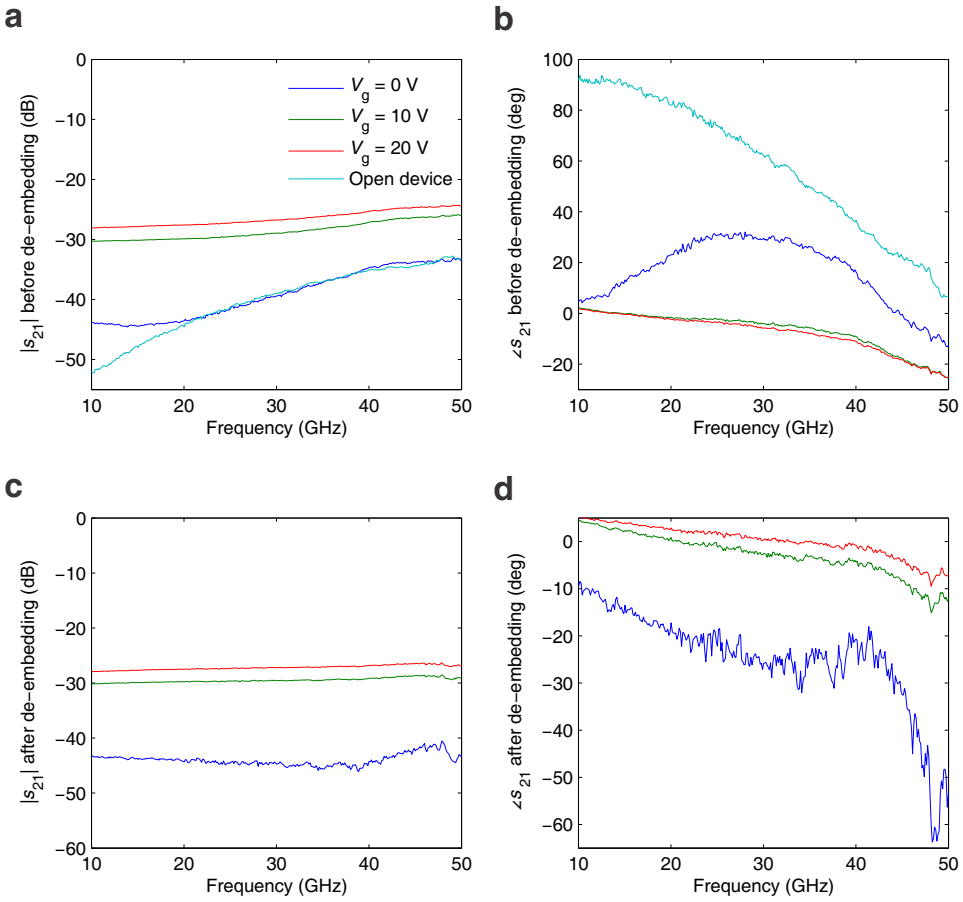

Figure 4.18: s_{21} measurements of ungated graphene device [16]. **a**, $|s_{21}|$ before parasitic signal de-embedding. **b**, $\angle s_{21}$ before de-embedding. **c**, $|s_{21}|$ after de-embedding. **d**, $\angle s_{21}$ after de-embedding.

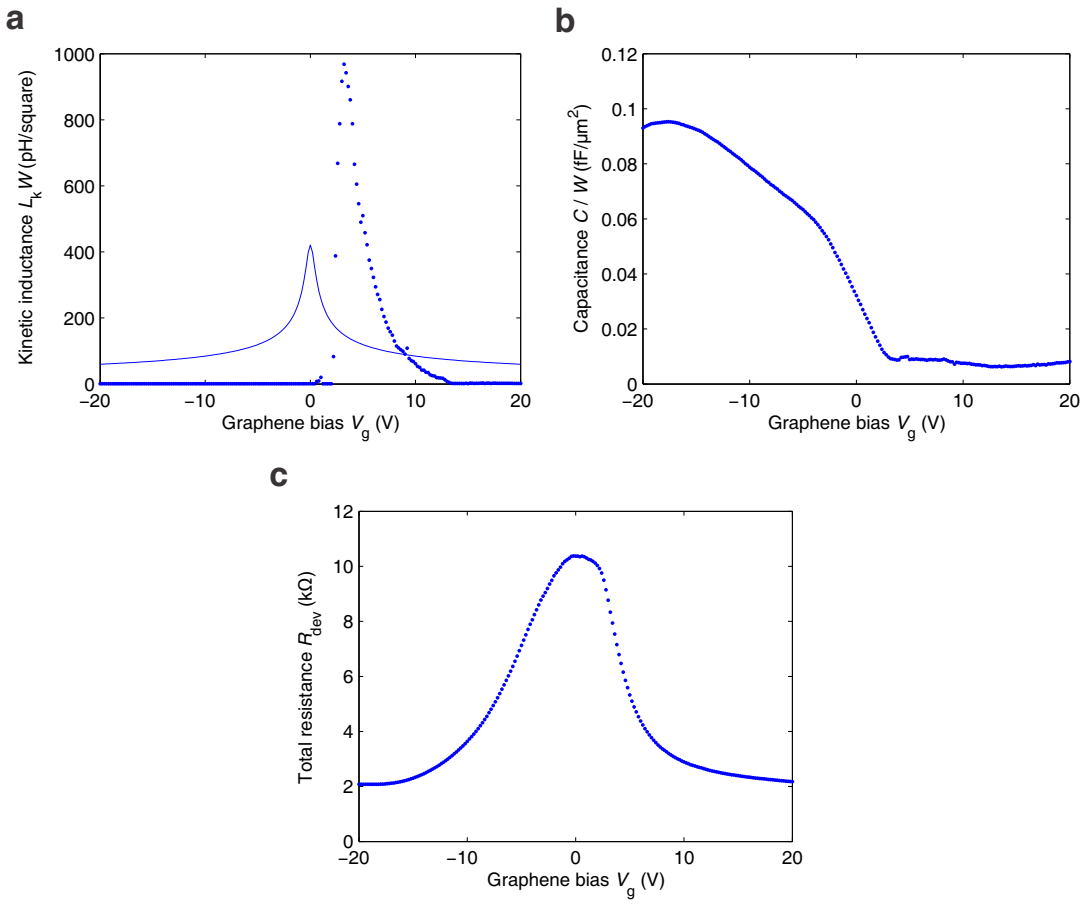

Figure 4.19: $L_{\rm k}$ (**a**), C (**b**), and $R_{\rm dev}$ (**c**) extracted for the ungated graphene device [16]. Solid curve in **a** is the theoretical prediction. This measurement was performed at a considerably later time compared to the DC measurement of Fig. 4.17, and shows a conspicuous shift of the charge neutrality point.

substantial change in the graphene bias (red and green curves, Fig. 4.10d) that must cause an appreciable change in $L_{\rm k}$ leads only to a small $\angle s_{21}$ difference of only a few degrees at best, further hampering $L_{\rm k}$ extraction. Device parameter extracted from these s-parameters are highly spurious (Fig. 4.19), with the final curve fits for s-parameters plagued with large residual errors.

4.1.4 Appendix: Materials and Methods

We fabricated h-BN encapsulated graphene by mechanical exfoliation and polymer-free mechanical transfer of h-BN single crystals and graphene through optical alignment [81]. High resistivity ($> 5000\ \Omega{\rm cm}$) silicon wafers coated with 285-nm thick thermal oxide were used as the substrate to minimize high-frequency substrate losses. Optical differentiation and Raman spectroscopy were used to confirm that graphene is single layered. Contacts [81] and waveguides were created by thermal evaporation of Cr/Pd/Au (1/10/300 nm) with dimensions defined by electron beam lithography and inductively coupled plasma etching.

Measurements took place in a Lake Shore Cryotronics cryogenic probe station at feedback-controlled temperatures in the dark. DC resistance measurements were performed using a Stanford Research Systems SR830 lock-in amplifier and a DL Instruments 1211 current preamplifier. Microwave s-parameter measurements were performed using an Agilent E8364A vector network analyzer, where the calibration was performed using the NIST-style multiline TRL technique [70] at each temperature just before the measurement. The parasitic coupling bypassing the graphene device was measured on a separate device with the identical CPW structures but with no h-BN encapsulated graphene, and was then de-embedded from the measured s-parameters of the main device [14] (see Appendix A for details).

The design of the CPWs was performed using a Sonnet frequency-domain electromagnetic field solver. The CPW dimensions were chosen to match the 50-Ω characteristic impedance of the network analyzer, cables, and probes [14].

4.2 Fluctuation and Collective Dynamics

In this section[10], we theoretically study a microscopic formulation of thermal current fluctuation, or Johnson-Nyquist noise, in graphene. While a conventional massive ($m \neq 0$) electron gas exhibits equipartition form of fluctuation that increases linearly with temperature, it is shown that the current fluctuation in graphene exhibits a more complicated dependency on temperature due to its unique band structure. Nonetheless, the total collective current fluctuation is shown to obey the equipartition form together with the collective mass or kinetic inductance of graphene electrons, implying that the collective current's degree of freedom follows a massive equation of motion. Hence, we show an intimate link between the fluctuation behaviors and the collective dynamics (including plasmonics) in graphene. This observation can also be generalized to an arbitrary electron gas.

4.2.1 Introduction

Thermal agitation of electrons in a conductor creates spontaneous current fluctuations, or Johnson-Nyquist noise [88, 89], whose power spectral density is $S_I(f) = 4k_{\mathrm{B}}TG$ (k_B: Boltzmann constant; T: temperature; G: conductance). Nyquist explained this formula with a *macroscopic* thought experiment [89], where the thermal noise energy coupled to electromagnetic modes of an external transmission line was considered with the equipartition theorem. The Johnson-Nyquist noise can also be explained without considering external entities, but by directly considering the thermal motions of electrons. In this *microscopic* approach [90], electrons (mass: m) are treated as classical particles following the Maxwell-Boltzmann distribution, and the mean squared thermal fluctuation velocity v_{f} ('f' notes fluctuation) of each electron is set by the equipartition theorem: $\langle v_{\mathrm{f}}^2 \rangle = k_{\mathrm{B}}T/m \times \mathrm{dimensions}$. The aggregate of this electron velocity fluctuation causes the total current fluctuation $\langle I^2 \rangle \propto T$, from which $S_I(f) = 4k_{\mathrm{B}}TG$ follows.

It is then of interest to study the microscopic machinery behind the thermal noise in graphene. Graphene's unique band structure causes a broad palette of surprising phenomena, for which the thermal stochastic dynamics is not expected to be an exception. Concretely, as individual graphene electrons act as massless relativistic particles [71], the equipartition theorem used in the traditional microscopic

[10] Large portions of this section are derived from a paper in publication by the author [17].

approach cannot be applied and $\langle I^2 \rangle \propto T$ will *not* hold. Moreover, the coexistence of electrons and holes due to graphene's zero-bandgap nature [71] can further enrich their fluctuation dynamics.

We will elucidate the unique thermal fluctuation behaviors of massless electrons (and holes) in graphene, *i.e.*, the nonlinear T-dependence of $\langle v_{\rm f}^2 \rangle$ and $\langle I^2 \rangle$, using a microscopic formalism based on Fermi-Dirac statistics, as the traditional approach with Maxwell-Boltzmann statistics is fundamentally limited. Despite the nonlinear T-dependence of $\langle I^2 \rangle$ in graphene, the fluctuation-dissipation theorem [91] still demands $S_I(f) = 4k_{\rm B}TG$ [92, 93]; showing how this occurs will be a corollary.

While individual graphene electrons are massless, their collective excitation exhibits a Newtonian inertia, which is essential for graphene plasmonics and has been measured recently [16]. We will unveil that the thermal fluctuation dynamics has an intimate connection to this collective mass; *i.e.*, $\langle I^2 \rangle$—while arising from massless electrons—is given by the equipartition theorem applied to the collective mass's kinetic energy, with the nonlinear T-dependence of $\langle I^2 \rangle$ arising from the collective mass's T-dependence. Moreover, we will show that this fluctuation-collective-dynamics relation manifested in $\langle I^2 \rangle$ is so general that it applies to any conductor beyond graphene.

4.2.2 Microscopic Formulation of Thermal Fluctuations

To this end, we first formulate the thermal fluctuation of electron velocity $\langle v_{\rm f}^2 \rangle$ and the corresponding total current fluctuation $\langle I^2 \rangle$ in a general conductor using Fermi-Dirac statistics. This formulation is applicable to conductors in any dimensions, but for simplicity, we consider a two-dimensional (2D) conductor, whether it be graphene with massless electrons or 2D conductors with massive ($m \neq 0$) electrons (*e.g.*, GaAs/AlGaAs quantum wells). An electron with a wavevector $\mathbf{k}$ assumes an intrinsic velocity of $v_{\mathbf{k}}$: for a massive 2D electron gas, $v_{\mathbf{k}} = \hbar k/m$, where $k \equiv |\mathbf{k}| = (k_x^2 + k_y^2)^{1/2}$; for graphene with massless electrons, $v_{\mathbf{k}} = v_{\rm F}$ (constant). These intrinsic velocities, even existing at $T = 0$, by themselves are not thermal velocity fluctuations but arise due purely to the quantum nature of the electron gas. The thermal velocity fluctuation $v_{\rm f}$ can be evaluated properly only if these intrinsic velocities are considered judiciously together with the Fermi-Dirac thermal distribution of electrons, $f_{\mathbf{k}} = 1/[e^{(\varepsilon_{\mathbf{k}}-\mu)/k_{\rm B}T} + 1]$ ($\varepsilon_{\mathbf{k}}$: single electron energy; μ: chemical potential).

Note first that $\langle v_{\rm f}^2 \rangle$ must *not* be the average of $v_{\mathbf{k}}^2$ across all electrons, $(1/n) \int (d^2\mathbf{k}/(2\pi)^2) g v_{\mathbf{k}}^2 f_{\mathbf{k}}$ (g: spin/valley degeneracy; n: electron density). This all-electron average includes electrons moving

in opposite directions with the same velocity, whose velocities cancel and cannot contribute to current fluctuations. Concretely, if a $\mathbf{k}$-state and a $-\mathbf{k}$-state sit deep below the Fermi surface with $f_{\mathbf{k}} = f_{-\mathbf{k}} = 1$, each of these states is surely occupied by an electron, and these two electrons do not contribute to current fluctuation; yet the all-electron average counts many such non-contributing pairs. Also note that the all-electron average does not vanish at $T = 0$, whereas $\langle v_{\mathrm{f}}^2 \rangle$ must at $T = 0$.

To evaluate $\langle v_{\mathrm{f}}^2 \rangle$, we only consider electrons whose velocities are not canceled. The probability that a $\mathbf{k}$-state is occupied and a $-\mathbf{k}$-state is *not* occupied is $f_{\mathbf{k}}(1 - f_{-\mathbf{k}})$, and thus $\langle v_{\mathrm{f}}^2 \rangle$ can be calculated as

$$\langle v_{\mathrm{f}}^2 \rangle = \frac{1}{n} \int \frac{d^2\mathbf{k}}{(2\pi)^2} g v_{\mathbf{k}}^2 f_{\mathbf{k}} (1 - f_{-\mathbf{k}}), \tag{4.24}$$

where the electron density n is

$$n = \int \frac{d^2\mathbf{k}}{(2\pi)^2} g f_{\mathbf{k}}. \tag{4.25}$$

Assuming $\varepsilon_{\mathbf{k}} = \varepsilon_{-\mathbf{k}}$, $f_{\mathbf{k}}(1 - f_{-\mathbf{k}})$ can be rewritten as

$$f_{\mathbf{k}}(1 - f_{-\mathbf{k}}) = \frac{\partial f_{\mathbf{k}}}{\partial(\mu/k_{\mathrm{B}}T)} = -\frac{\partial f_{\mathbf{k}}}{\partial(\varepsilon_{\mathbf{k}}/k_{\mathrm{B}}T)}, \tag{4.26}$$

which we will make use of later. At low T, since $f_{\mathbf{k}}(1 - f_{-\mathbf{k}})$ in $\mathbf{k}$-space peaks around the Fermi surface with a vanishing width for $T \to 0$, $\langle v_{\mathrm{f}}^2 \rangle$ vanishes at $T = 0$, as it should. Note that, for graphene, while each electron has the constant velocity ($v_{\mathbf{k}} = v_{\mathrm{F}}$), Eq. (4.24) shows that $\langle v_{\mathrm{f}}^2 \rangle \neq v_{\mathrm{F}}^2$, as $\langle v_{\mathrm{f}}^2 \rangle$ is determined not solely by $v_{\mathbf{k}} = v_{\mathrm{F}}$, but also by the probability distribution $f_{\mathbf{k}}(1 - f_{-\mathbf{k}})$. We will calculate $\langle v_{\mathrm{f}}^2 \rangle$ for graphene shortly.

$\langle v_{\mathrm{f}}^2 \rangle$ leads to the total current thermal fluctuation $\langle I^2 \rangle$. Consider a 2D conductor strip of width W and length l along the x axis, and let $\langle I^2 \rangle$ be measured along the length. Then only the x-component of v_{f}, or $v_{\mathrm{f},x}$, contributes to the overall current fluctuation. Since a single electron would contribute a fluctuation current of $(e v_{\mathrm{f},x}/l)^2$, and since there are a total of nWl electrons,

$$\langle I^2 \rangle = nWl \frac{e^2}{l^2} \langle v_{\mathrm{f},x}^2 \rangle = ne^2 \frac{W}{l} \langle v_{\mathrm{f},x}^2 \rangle, \tag{4.27}$$

where $\langle v_{\mathrm{f},x}^2 \rangle = \langle v_{\mathrm{f}}^2 \rangle / 2$ follows from angular integrations.

$S_I(f)$ readily follows from $\langle I^2 \rangle$. The autocorrelation of the stationary process I is [94] $\langle I(0)I(t) \rangle =$

$\langle I^2\rangle e^{-|t|/\tau}$ (τ: Drude scattering time), because electron scatterings randomize initial momenta at an average rate of $1/\tau$. It is in this step that the relation between the thermal fluctuation and the dissipative behavior is established. The single-sided power spectral density is then $S_I(f) = 4\int_0^\infty dt\langle I(0)I(t)\rangle \cos(\omega t)$ with $\omega = 2\pi f$, or,

$$S_I(f) = 4\langle I^2\rangle \frac{\tau}{1+\omega^2\tau^2}. \tag{4.28}$$

To check the cogency of this formalism, we first apply it to a massive 2D electron gas, as the result can be compared to the traditional approach [90] valid for massive electron gas. Using $\varepsilon_{\mathbf{k}} = \hbar^2k^2/2m$, $v_{\mathbf{k}} = \hbar k/m$, Eqs. (4.25) and (4.26), and $\langle v_{\mathrm{f},x}^2\rangle = \langle v_{\mathrm{f}}^2\rangle/2$ in Eq. (4.24), we find

$$\langle v_{\mathrm{f},x}^2\rangle = \frac{k_{\mathrm{B}}T}{m}\frac{\int_0^\infty d\xi\xi\frac{\partial}{\partial\eta}f(\xi-\eta)}{\int_0^\infty d\xi f(\xi-\eta)}, \tag{4.29}$$

where $\xi \equiv \varepsilon_k/k_{\mathrm{B}}T$, $\eta \equiv \mu/k_{\mathrm{B}}T$, and $f(\xi) \equiv 1/(e^\xi+1)$. Using $\int_0^\infty d\xi\xi^s f(\xi-\eta) = -\Gamma(1+s)\mathrm{Li}_{1+s}(-e^\eta)$, where $\Gamma(z)$ is the gamma function and $\mathrm{Li}_n(z) = \sum_{k=1}^\infty z^k/k^n$ is the polylogarithm function, we reduce Eq. (4.29) to

$$\langle v_{\mathrm{f},x}^2\rangle = \frac{k_{\mathrm{B}}T}{m}\frac{\frac{\partial}{\partial\eta}\mathrm{Li}_2(-e^\eta)}{\mathrm{Li}_1(-e^\eta)} = \frac{k_{\mathrm{B}}T}{m}, \tag{4.30}$$

where we have used $\frac{d}{dx}\mathrm{Li}_n(x) = \frac{1}{x}\mathrm{Li}_{n-1}(x)$. This is consistent with the traditional approach [90] based on Maxwell-Boltzmann statistics, where Eq. (4.30) results from the equipartition theorem. Eq. (4.27) then yields

$$\langle I^2\rangle = \frac{ne^2}{m}\frac{W}{l}k_{\mathrm{B}}T. \tag{4.31}$$

In sum, for a massive electron gas, our general approach and the traditional approach agree; importantly, $\langle v_{\mathrm{f},x}^2\rangle \propto T$ and $\langle I^2\rangle \propto T$. Incidentally, Eq. (4.28) then yields $S_I(f) = 4k_{\mathrm{B}}T[(ne^2\tau/m)(1+\omega^2\tau^2)^{-1}](W/l)$, where the real part of the Drude conductivity $\sigma = (ne^2\tau/m)/(1+i\omega\tau)$ appears inside the square brackets. As the real part of $\sigma W/l$ is G, we arrive at $S_I(f) = 4k_{\mathrm{B}}TG$.

4.2.3 Fluctuations in Graphene

We now apply the formalism to graphene with $\varepsilon_{\mathbf{k}} = \pm\hbar v_{\mathrm{F}}k$ and $v_{\mathbf{k}} = v_{\mathrm{F}}$ [3]. This constant $v_{\mathbf{k}}$ of massless individual electrons and holes will lead to a nonlinear T-dependency of $\langle v_{\mathrm{f},x}^2\rangle$ and $\langle I^2\rangle$, sharply

Table 4.1: T-dependence of μ and η for a fictitious graphene with conduction or valence band only (Fig. 4.20), or for the actual graphene with both bands (Figs. 4.21 and 4.22) [17].

		$T=0$		$T\to\infty$	
Band	Constraint	μ	η	μ	η
Conduction only	$n_{\rm e}$	$\varepsilon_{\rm F}$	$+\infty$	$-\infty$	$-\infty$
Valence only	$n_{\rm h}$	$-\varepsilon_{\rm F}$	$-\infty$	$+\infty$	$+\infty$
Conduction and valence	$n_{\rm e}-n_{\rm h}$	$\pm\varepsilon_{\rm F}$	$\pm\infty$	0	0

contrasting the linear T-dependency of the massive case. The coexistence of electrons and holes further enriches the thermal fluctuation behaviors. For graphene, $\langle v_{{\rm f},x}^2\rangle$ and n are calculated separately for electrons in the conduction band and for holes in the valence band ('e': electrons, 'h': holes):

$$\langle v_{{\rm ef},x}^2\rangle = \frac{v_{\rm F}^2}{2}\frac{{\rm Li}_1(-e^{\eta})}{{\rm Li}_2(-e^{\eta})},\quad \langle v_{{\rm hf},x}^2\rangle = \frac{v_{\rm F}^2}{2}\frac{{\rm Li}_1(-e^{-\eta})}{{\rm Li}_2(-e^{-\eta})}, \tag{4.32}$$

$$n_{\rm e} = \frac{-g(k_{\rm B}T)^2}{2\pi(\hbar v_{\rm F})^2}{\rm Li}_2(-e^{\eta}),\quad n_{\rm h} = \frac{-g(k_{\rm B}T)^2}{2\pi(\hbar v_{\rm F})^2}{\rm Li}_2(-e^{-\eta}). \tag{4.33}$$

Here, $f_{\mathbf{k}} = f(\xi+\eta)$ and $f_{\mathbf{k}}(1-f_{-\mathbf{k}}) = -(\partial/\partial\eta)f_{\mathbf{k}} = -(\partial/\partial\xi)f_{\mathbf{k}}$ are used for the distribution of holes.

To first see the massless effect without the complication from the electron-hole coexistence, consider a fictitious graphene with the conduction band only (electrons only) with $\varepsilon_{\mathbf{k}} = \hbar v_{\rm F}k$. The T-dependency depends on whether the chemical potential μ or electron density $n_{\rm e}$ is fixed for varying T. We consider the constant $n_{\rm e}$ case, as it is practically achieved with electrostatic gating. Then $n_{\rm e} = {\rm constant}$ condition [Eq. (4.33)] determines $\mu(T)$ with $\mu(0) = \varepsilon_{\rm F} = \hbar v_{\rm F}\sqrt{4\pi n_{\rm e}/g} > 0$ [Table 4.1]. With this particular $\mu(T)$, $\langle v_{{\rm ef},x}^2\rangle$ first grows linearly with T just as in the massive case, but eventually saturates to $v_{\rm F}^2/2$, drastically deviating from the persistent linear T-dependence of the massive case [Fig. 4.20].

This low-T similarity, high-T difference between the massless and massive case can be explained with Eq. (4.24). For $k_{\rm B}T \ll \varepsilon_{\rm F}$, $f_{\mathbf{k}}(1-f_{-\mathbf{k}})$ peaks sharply around the Fermi surface, so $v_{\mathbf{k}} = v_{\rm F}$ for graphene coincides with $v_{\mathbf{k}} \approx v_{\rm F}$ for the massive case, while this peak's width grows linearly with T. So Eq. (4.24) is linear to T in both massless and massive cases. For $k_{\rm B}T \gg \varepsilon_{\rm F}$ with $\mu\to-\infty$ [Table 4.1], in the conduction band, $f_{\mathbf{k}}(1-f_{-\mathbf{k}}) \approx f_{\mathbf{k}} \approx e^{-(\varepsilon_{\mathbf{k}}-\mu)/k_{\rm B}T}$ is the far tail of the Fermi-Dirac distribution. So $v_{\mathbf{k}}^2 = v_{\rm F}^2$ (massless) and $v_{\mathbf{k}}^2 \propto k^2$ (massive) makes a difference in Eq. (4.24); in the former, $\langle v_{{\rm ef},x}^2\rangle$

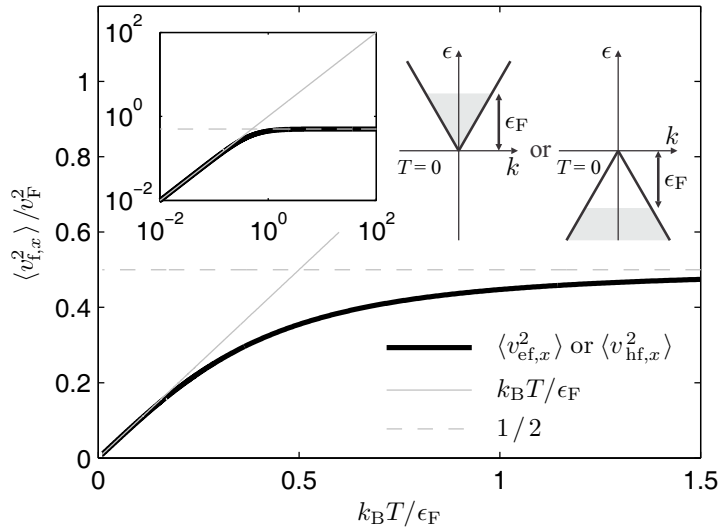

Figure 4.20: T-dependence of $\langle v^2_{\mathrm{f},x}\rangle$ for fictitious conduction-band-only graphene with constant n_{e}, or valence-band-only graphene with constant n_{h} [17]. Inset: the same plot, log scales.

saturates; in the latter, $\langle v^2_{\mathrm{ef},x}\rangle \propto T$ persists.

We can also consider a fictitious graphene with the valence band only (holes only) with $\varepsilon_{\mathbf{k}} = -\hbar v_{\mathrm{F}} k$. In this case, $n_{\mathrm{h}} = \text{constant}$ [Eq. (4.33)] determines $\mu(T)$ with $\mu(0) = -\varepsilon_{\mathrm{F}} < 0$ [Table 4.1]. The resulting T-dependence of $\langle v^2_{\mathrm{hf},x}\rangle$ is exactly the same as that of $\langle v^2_{\mathrm{ef},x}\rangle$ [Fig. 4.20].

Now consider the actual graphene with both the conduction and valence bands. Let graphene be electron-doped at $T = 0$ and the total charge density $\propto n_{\mathrm{e}}(T) - n_{\mathrm{h}}(T)$ be kept constant via electrostatic gating. $\mu(0) = \varepsilon_{\mathrm{F}} = \hbar v_{\mathrm{F}}\sqrt{4\pi n_{\mathrm{e}}(0)/g} > 0$, $n_{\mathrm{h}}(0) = 0$, and $n_{\mathrm{e}}(T) - n_{\mathrm{h}}(T) = n_{\mathrm{e}}(0)$ for any T. Using $n_{\mathrm{e}}(T)$ and $n_{\mathrm{h}}(T)$ from Eq. (4.33), this last expression can be rewritten as

$$\frac{g(k_{\mathrm{B}}T)^2}{2\pi(\hbar v_{\mathrm{F}})^2}[-\mathrm{Li}_2(-e^{\eta}) + \mathrm{Li}_2(-e^{-\eta})] = n_{\mathrm{e}}(0). \tag{4.34}$$

Eq. (4.34) determines $\mu(T)$ [Table 4.1]. $\mu = 0$ for $T \to \infty$, which contrasts the electron- or hole-only case, because n_{e} and n_{h} grow increasingly similar with T ($n_{\mathrm{e}}/n_{\mathrm{h}} \to 1$) despite their fixed difference. $\langle v^2_{\mathrm{ef},x}\rangle$ and $\langle v^2_{\mathrm{hf},x}\rangle$ are still given by Eq. (4.32), but due to the new $\mu(T)$, T-dependence of $\langle v^2_{\mathrm{ef},x}\rangle$ and $\langle v^2_{\mathrm{hf},x}\rangle$ [Fig. 4.21] now deviates from Fig. 4.20.

$\langle v^2_{\mathrm{ef},x}\rangle$ is still linear to small T, as the actual electron-doped graphene in this regime is no different from the fictitious, electron-only graphene. For $T \to \infty$, $\langle v^2_{\mathrm{ef},x}\rangle$ also saturates, but to $(6\ln(2)/\pi^2)v^2_{\mathrm{F}}$ instead of $v^2_{\mathrm{F}}/2$, because $\mu(T \to \infty) = 0$ now, while $\mu(T \to \infty) \to -\infty$ in the electron-only graphene.

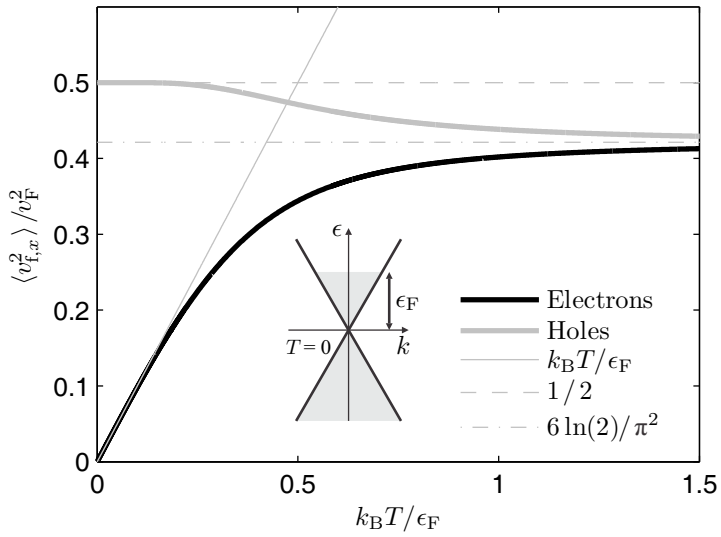

Figure 4.21: T-dependence of $\langle v_{\mathrm{ef},x}^2 \rangle$ and $\langle v_{\mathrm{hf},x}^2 \rangle$ for electron-doped graphene ($\mu(0) > 0$) with $\varepsilon_{\mathbf{k}} = \pm\hbar v_{\mathrm{F}} k$, assuming constant charge density (*i.e.*, $n_{\mathrm{e}} - n_{\mathrm{h}} = \text{constant}$) [17].

$\langle v_{\mathrm{hf},x}^2 \rangle$ in Fig. 4.21 more drastically differs from Fig. 4.20, as we start from an electron-doped graphene. The small number of holes in the valence band at low T are at the far tail of the Fermi-Dirac distribution (similar to the $T \to \infty$ case of Fig. 4.20), so $\langle v_{\mathrm{hf},x}^2 \rangle \to v_{\mathrm{F}}^2/2$ for low T. For large T, $\mu \to 0$, so $\langle v_{\mathrm{hf},x}^2 \rangle$ approaches $(6\ln(2)/\pi^2)v_{\mathrm{F}}^2$ just like $\langle v_{\mathrm{ef},x}^2 \rangle$. In a similar manner, the T-dependencies for a hole-doped graphene are exactly the same as seen above but with the roles of electrons and holes reversed.

These unique behaviors of $\langle v_{\mathrm{ef},x}^2 \rangle$ and $\langle v_{\mathrm{hf},x}^2 \rangle$ yield a unique T-dependence of $\langle I^2 \rangle$. As electron and hole current fluctuations are independent, Eq. (4.27) is now $\langle I^2 \rangle = (We^2/l)[n_{\mathrm{e}}\langle v_{\mathrm{ef,x}}^2 \rangle + n_{\mathrm{h}}\langle v_{\mathrm{hf,x}}^2 \rangle]$, or

$$\langle I^2 \rangle = \frac{ge^2 W}{4\pi\hbar^2 l}(k_{\mathrm{B}}T)^2[-\mathrm{Li}_1(-e^{\eta}) - \mathrm{Li}_1(-e^{-\eta})], \tag{4.35}$$

using Eqs. (4.32) and (4.33). Figure 4.22 plots $\langle I^2 \rangle$ vs. T with $\mu(T)$ set by Eq. (4.34). At low T, as $n_{\mathrm{e}}(T) \approx n_{\mathrm{e}}(0)$, $n_{\mathrm{h}}(T) \approx 0$, and $\langle v_{\mathrm{ef,x}}^2 \rangle \propto T$, we have $\langle I^2 \rangle \propto \langle v_{\mathrm{ef,x}}^2 \rangle \propto T$ (or electrons/holes reversed for a hole-doped graphene). At high T, as both $\langle v_{\mathrm{ef},x}^2 \rangle$ and $\langle v_{\mathrm{hf},x}^2 \rangle$ converge to $(6\ln(2)/\pi^2)v_{\mathrm{F}}^2$, and as both n_{e} and n_{h} grow with T^2 (see Eq. (4.34) with $\mu = 0$ for $T \to \infty$), $\langle I^2 \rangle \propto (n_{\mathrm{e}} + n_{\mathrm{h}}) \propto T^2$. In sum, the massless-ness of electrons and holes and their coexistence in graphene yield unique thermal fluctuation dynamics. In particular, $\langle v_{\mathrm{ef},x}^2 \rangle$, $\langle v_{\mathrm{hf},x}^2 \rangle$, and $\langle I^2 \rangle$ vary nonlinearly with T, contrasting the linear T-dependence in massive electron gases.

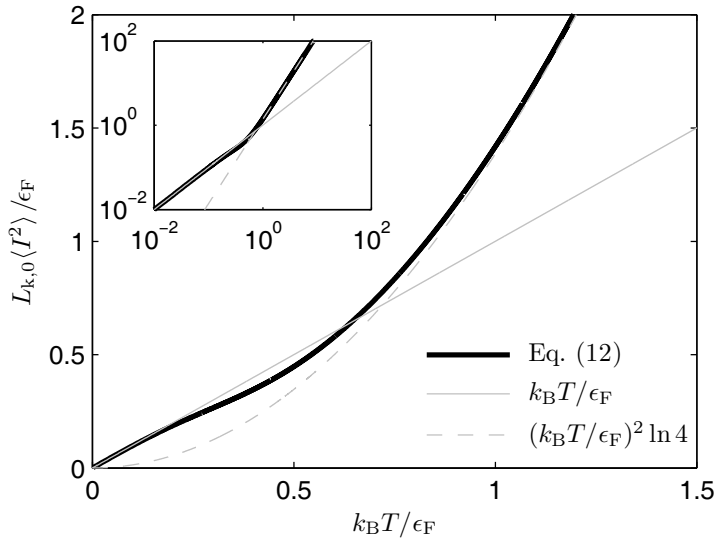

Figure 4.22: $\langle I^2\rangle$ vs. T for electron- or hole- doped graphene with constant charge density (*i.e.*, $n_{\rm e} - n_{\rm h} = \text{constant}$); $L_{\rm k,0} = 4\pi\hbar^2 l/ge^2\varepsilon_{\rm F}W$ [17]. Inset: same plot, log scales.

Incidentally, graphene intraband conductivity is [31]

$$\sigma = \frac{-ge^2 k_{\rm B}T}{4\pi\hbar^2(\tau^{-1}+i\omega)}\int_0^\infty d\xi\xi\Big(\frac{\partial f(\xi-\eta)}{\partial\xi} - \frac{\partial f(-\xi-\eta)}{\partial\xi}\Big)$$

$$= \frac{ge^2 k_{\rm B}T}{4\pi\hbar^2(\tau^{-1}+i\omega)}[-{\rm Li}_1(-e^{\eta}) - {\rm Li}_1(-e^{-\eta})], \tag{4.36}$$

where the conduction and valence band contributions are separated. Comparing the real part of the above with Eq. (4.35) and noting that ${\rm Re}[\sigma W/l] = G$, we attain $\langle I^2\rangle = k_{\rm B}TG(1+\omega^2\tau^2)/\tau$. By plugging this into Eq. (4.28), we arrive at $S_I(f) = 4k_{\rm B}TG$. Despite the distinct T-dependence of $\langle I^2\rangle$ in graphene, as σ shows the same T-dependence except for the $k_{\rm B}T$ factor, the Johnson-Nyquist noise expression still holds. This is how the fluctuation-dissipation relation [91] manifests in graphene.

4.2.4 Fluctuation and Collective Dynamics

We now make a fundamental observation on the intimate relation between the thermal fluctuation $\langle I^2\rangle$ and collective dynamics (first for graphene, then generally). To this end, we first briefly discuss the collective motion of graphene electrons [16], while setting aside the fluctuation. Imagine graphene electrons collectively moving due to a voltage V. Individual electrons' velocity $v_{\rm F}$ remains constant, but their wavevectors change by $\Delta\hat{\mathbf{x}}$. The total kinetic energy of the electron gas has then grown by an

amount $E_{\rm e}$. It can be shown that $E_{\rm e}$ assumes a smooth minimum at $\Delta = 0$ for graphene, *i.e.*, $E_{\rm e} \propto \Delta^2$ for small enough Δ [16]. On the other hand, the collective momentum $P_{\rm e} \propto \Delta$. Hence, $E_{\rm e} \propto P_{\rm e}^2$ and the collective motion exhibits a Newtonian inertia $M_{\rm e}$ satisfying $E_{\rm e} = P_{\rm e}^2/(2M_{\rm e})$, while individual electrons are massless.

Thus $M_{\rm e}$ accelerates according to the Newton's law, increasing its velocity $V_{\rm ec} \equiv P_{\rm e}/M_{\rm e}$. The frequency-domain equation of motion is $-(n_{\rm e}Wl)(eV/l) = i\omega M_{\rm e}V_{\rm ec}$. As the current is $I_{\rm e} = -n_{\rm e}eWV_{\rm ec}$, $V/I_{\rm e} = i\omega[M_{\rm e}/(en_{\rm e}W)^2] = i\omega L_{\rm ek}$, yielding the kinetic inductance of the electron gas. Noting that the same holds true for holes,

$$L_{\rm ek} = (e^2 n_{\rm e}^2 W^2)^{-1} M_{\rm e}, \quad L_{\rm hk} = (e^2 n_{\rm h}^2 W^2)^{-1} M_{\rm h}. \tag{4.37}$$

That is, the collective acceleration can be described using $M_{\rm e}$ ($M_{\rm h}$) or intimately related $L_{\rm ek}$ ($L_{\rm hk}$). Note that $E_{\rm e} = M_{\rm e}V_{\rm ec}^2/2 = L_{\rm ek}I_{\rm e}^2/2$ and $E_{\rm h} = M_{\rm h}V_{\rm hc}^2/2 = L_{\rm hk}I_{\rm h}^2/2$.

We can find the expressions of $L_{\rm ek}$ and $L_{\rm hk}$ in graphene from Eq. (4.36). As $\omega L_{\rm k} = {\rm Im}[l/\sigma W]$, we have

$$L_{\rm k} = \frac{4\pi\hbar^2}{ge^2 k_{\rm B}T} \frac{1}{[-{\rm Li}_1(-e^{\eta}) - {\rm Li}_1(-e^{-\eta})]} \frac{l}{W}. \tag{4.38}$$

This is the overall kinetic inductance combining $L_{\rm ek}$ and $L_{\rm hk}$ in parallel as $L_{\rm k}^{-1} = L_{\rm ek}^{-1} + L_{\rm hk}^{-1}$, with

$$L_{\rm ek} = \frac{-4\pi\hbar^2}{ge^2 k_{\rm B}T} \frac{l/W}{{\rm Li}_1(-e^{\eta})}, \quad L_{\rm hk} = \frac{-4\pi\hbar^2}{ge^2 k_{\rm B}T} \frac{l/W}{{\rm Li}_1(-e^{-\eta})}. \tag{4.39}$$

In the collective dynamics discussed above, we set aside the fluctuation problem. In fact, graphene plasmonic waves—a non-fluctuation (non-stochastic) event—can occur owing to the kinetic inductance and collective mass, which have been recently measured in graphene [16]. We now turn back to the fluctuation problem and find its deep-seated connection to the collective dynamics.

Comparing the total current fluctuation $\langle I^2 \rangle$ [Eq. (4.35)] to the total kinetic inductance $L_{\rm k}$ [Eq. (4.38)] yields

$$\frac{1}{2} L_{\rm k} \langle I^2 \rangle = \frac{1}{2} k_{\rm B} T. \tag{4.40}$$

This can be broken into electron and hole contributions:

$$\frac{1}{2}L_{\rm ek}\langle I_{\rm e}^2\rangle = \frac{1}{2}k_{\rm B}T, \quad \frac{1}{2}L_{\rm hk}\langle I_{\rm h}^2\rangle = \frac{1}{2}k_{\rm B}T, \tag{4.41}$$

as $\langle I^2\rangle = \langle I_{\rm e}^2\rangle + \langle I_{\rm h}^2\rangle$ and $L_{\rm k}^{-1} = L_{\rm ek}^{-1} + L_{\rm hk}^{-1}$. Or equivalently, in terms of $M_{\rm e}$ and $M_{\rm h}$, and their thermal velocity fluctuations $\langle V_{\rm ec}^2\rangle$ and $\langle V_{\rm hc}^2\rangle$,

$$\frac{1}{2}M_{\rm e}\langle V_{\rm ec}^2\rangle = \frac{1}{2}k_{\rm B}T, \quad \frac{1}{2}M_{\rm h}\langle V_{\rm hc}^2\rangle = \frac{1}{2}k_{\rm B}T. \tag{4.42}$$

Eqs. (4.40), (4.41), and (4.42) are the same statement on the intimate relation between thermal fluctuations and collective dynamics. Although individual graphene electrons and holes act as massless relativistic particles, their thermal fluctuations are governed by the classical kinetic energies of the collective electron mass $M_{\rm e}$ and of the collective hole mass $M_{\rm h}$, with each receiving a thermal energy of $k_{\rm B}T/2$, satisfying the equipartition theorem [Eq. (4.42)], thus determining the collective velocity thermal fluctuations $\langle V_{\rm ec}^2\rangle$ and $\langle V_{\rm ec}^2\rangle$. These directly translate to the thermal current fluctuations of electrons and holes, $\langle I_{\rm e}^2\rangle$ and $\langle I_{\rm h}^2\rangle$ [Eq. (4.41)]. Eq. (4.40) expresses this most concisely; the total current thermal fluctuation $\langle I^2\rangle$ is determined by the total kinetic inductance storing an average collective kinetic energy of $k_{\rm B}T/2$.

Eq. (4.40) holds even for the massive electron gas; Eq. (4.31) has the same form as Eq. (4.40), as $(m/ne^2)(l/W)$ is the kinetic inductance of massive electron gas, as well known from the Drude model. However, the energy equipartition for the collective kinetic energy in the massive case is less surprising, as each electron already follows equipartition and the collective mass is their simple aggregate ($M = nWlm$). In sum, in both massless and massive cases, $\langle I^2\rangle$ arises from the *unified* equipartition form of Eq. (4.40) applied to the collective kinetic energy stored in kinetic inductance (or collective mass). Importantly, despite the innocuous appearance of $\langle I^2\rangle = k_{\rm B}T/L_{\rm k}$ [Eq. (4.40)], in the massless case, L_k is decisively temperature dependent [Eq. (4.38)]; this is the key to the unique nonlinear T dependence of $\langle I^2\rangle$ in graphene. In the massive electron gas, L_k is independent of T, thus, $\langle I^2\rangle \propto T$.

In fact, Eq. (4.40), which we write here again—as the key result—with emphasis on the T-dependence of L_k,

$$\frac{1}{2}L_{\rm k}(T)\langle I^2\rangle = \frac{1}{2}k_{\rm B}T, \tag{4.43}$$

holds for *any* conductor regardless of the individual electron's energy dispersion (massive, massless, non-constant mass) and conductor dimensions. To demonstrate, we write the conductivity along the x-axis in a b-th band (note subscripts 'b' in the equations below) in a d-dimensional conductor with a general $\varepsilon_{\mathbf{k}}$ (or $v_{\mathbf{k}}$) [30]:

$$\sigma_b = \frac{e^2}{\tau_b^{-1} + i\omega} \int \frac{d^d\mathbf{k}}{(2\pi)^d} g v_{b\mathbf{k},x}^2 \Big(-\frac{\partial f_{\mathbf{k}}}{\partial \varepsilon_{b\mathbf{k}}}\Big). \tag{4.44}$$

In graphene, this reduced to either the electron or hole term in Eq. (4.36). Thus, the corresponding kinetic inductance $L_{b\mathrm{k}} = \omega^{-1}\mathrm{Im}[l/\sigma_b W]$ is given by:

$$L_{b\mathrm{k}} = \left[e^2 \int \frac{d^d\mathbf{k}}{(2\pi)^d} g v_{b\mathbf{k},x}^2 \Big(-\frac{\partial f_{\mathbf{k}}}{\partial \varepsilon_{b\mathbf{k}}}\Big)\right]^{-1} \frac{l}{W}. \tag{4.45}$$

On the other hand, the mean squared electron velocity fluctuation along the x-axis in the b-th band is

$$\langle v_{b\mathrm{f},x}^2 \rangle = \frac{k_{\mathrm{B}}T}{n_b} \int \frac{d^d\mathbf{k}}{(2\pi)^d} g v_{b\mathbf{k},x}^2 \Big(-\frac{\partial f_{\mathbf{k}}}{\partial \varepsilon_{b\mathbf{k}}}\Big), \tag{4.46}$$

where we have used Eqs. (4.24) and (4.26). Comparing the two equations above, we note $\langle v_{b\mathrm{f},x}^2 \rangle = (k_{\mathrm{B}}T/e^2 n_b)(1/L_{b\mathrm{k}})(l/W)$, then $\langle I^2 \rangle = \sum_b (e^2 n_b W/l)\langle v_{b\mathrm{f},x}^2 \rangle = k_{\mathrm{B}}T \sum_b (1/L_{b\mathrm{k}})$. But as the total kinetic inductance is $L_{\mathrm{k}} = [\sum_b (1/L_{b\mathrm{k}})]^{-1}$, we arrive at $\langle I^2 \rangle = k_{\mathrm{B}}T/L_{\mathrm{k}}$, which is Eq. (4.43).

In sum, regardless of the inner structure for individual electrons, their collective excitation exhibits a definite collective mass and corresponding kinetic inductance $L_{\mathrm{k}}(T)$, with thermal current fluctuation $\langle I^2 \rangle$ determined by applying the equipartition theorem to the collective kinetic energy stored in $L_{\mathrm{k}}(T)$ [Eq. (4.43)]. The detailed inner structure determines the specific T-dependency of $L_{\mathrm{k}}(T)$, and thus $\langle I^2 \rangle$ has a unique T-dependence reflecting the internal nature of the conductor.

As a final remark, we note that $\langle I^2 \rangle = \int_0^\infty df\, S_I(f)$ assumes a finite value regardless of τ, even in the lossless case with $\tau \to \infty$ (*i.e.*, $G = 0$) [Eq. (4.28)]; *cf.* $S_I(f) = 4k_{\mathrm{B}}TG$. That is, $\langle I^2 \rangle = k_{\mathrm{B}}T/L_{\mathrm{k}}$ holds with or without loss, as Eq. (4.46) makes no connection to electron scatterings. In this sense, $\langle I^2 \rangle = k_{\mathrm{B}}T/L_{\mathrm{k}}$ represents the thermal fluctuation in its very intrinsic form, stored in a self-reactive collective component L_k.

Eq. (4.43) also suggests a way to measure L_{k} via noise measurement. In low dimensional conductors, electromagnetic measurements of L_{k} have proven daunting, despite the importance of L_{k} in probing

plasmonic behaviors in general, and Luttinger liquid behaviors in one-dimensional conductors. In fact, L_{k} measurement in graphene has been only recently performed by the authors through a very delicate electromagnetic measurement [16]; and the decisive L_{k} measurement in one-dimensional conductors such as quantum wires and carbon nanotubes has yet to be carried out. The $\langle I^2 \rangle$ measurement and use of Eq. (4.43) might provide a new way of tackling this long-standing experimental problem.

Appendix A

Microwave Measurement and Analysis

Most of the microwave measurements appearing in this thesis were performed using Agilent E8364A two-port vector network analyzer and Cascade Microtech WinCal XE 4.2 software. Here we go over the details of the microwave measurement and analysis including calibration and de-embedding techniques for future reference. Large portion of the measurement and analysis techniques to be discussed here are shared in the three experimental works introduced in the thesis, with additional details pertaining to a specific work provided in its respective chapter.

Network Analyzer and WinCal XE Settings

Network analyzer front panel connections

PORT 1	**PORT 2**
CPLR ARM−SOURCE OUT	SOURCE OUT−CPLR ARM
RCVR A IN−CPLR THRU	CPLR THRU−RCVR B IN
REFERENCE 1	**REFERENCE 2**
SOURCE OUT−RCVR R1 IN	RCVR R2 IN−SOURCE OUT

Table A.1: 'Stimulus Settings' for WinCal XE.

Sweep Mode	Lin
Freq Segment 1	45 MHz ~ 50 GHz $(401$ points$)$
IF Bandwidth	30 Hz
Averaging	Off
Alternate Sweeps	Off
Interpolation	On
Port Extensions	Disable
Port Power Coupled	On
Port 1 Power	-30 dBm
Enable Power Slope	Enable
Power Slope	0.14 dB/GHz
Auto Attenuation	On
System z0	50 ohm

Calibration

Prior to any sets of measurements at a new temperature, Multiline TRL calibration [70] was performed. The calibration method involves landing microwave probes on thru/open/short structures and multiple sets of lines (CPWs) of different lengths (Fig. A.1; this mask file was created by W. F. Andress [11]).

Calibration structures were fabricated on undoped GaAs substrates using standard photolithography, with 8-nm Cr / 300-nm$\sim$500-nm Au. The widths of the signal line, ground lines, and the gaps between the two are determined using Sonnet simulations to match the 50-Ω characteristic impedance of the network analyzer, cables, and probes (see Sec. 3.2.4 for detailed simulation settings). For 100-μm pitch microwave probes, 24 μm/15 μm/123 μm (signal line/gap/ground line) widths were used, and for 150-μm pitch microwave probes, 50 μm/32 μm/ 143 μm (signal line/gap/ground line) widths were used.

After the calibration, the loss and phase delay from the microwave cables and probes are calibrated out, and the reference planes are set at the tips of the probes (Fig. A.2a). In addition, the r-l-g-c parameters (per-unit-length series resistance/series inductance/shunt conductance/shunt capacitance) of the CPWs in the calibration substrate (Fig. A.3) and the calibrated measurement results from the CPWs (total of 8 lines: 150 μm / 500 μm / 1000 μm / 1500 μm / 2200 μm / 3000 μm / 6000 μm / 9000 μm) are displayed, which can be used to determine if the calibration was sound.

Measurements for the actual devices are now performed, whose signal components contain the delays in the CPWs that lead to [from] the device-under-test (DUT) from [to] the probe tips, the response of

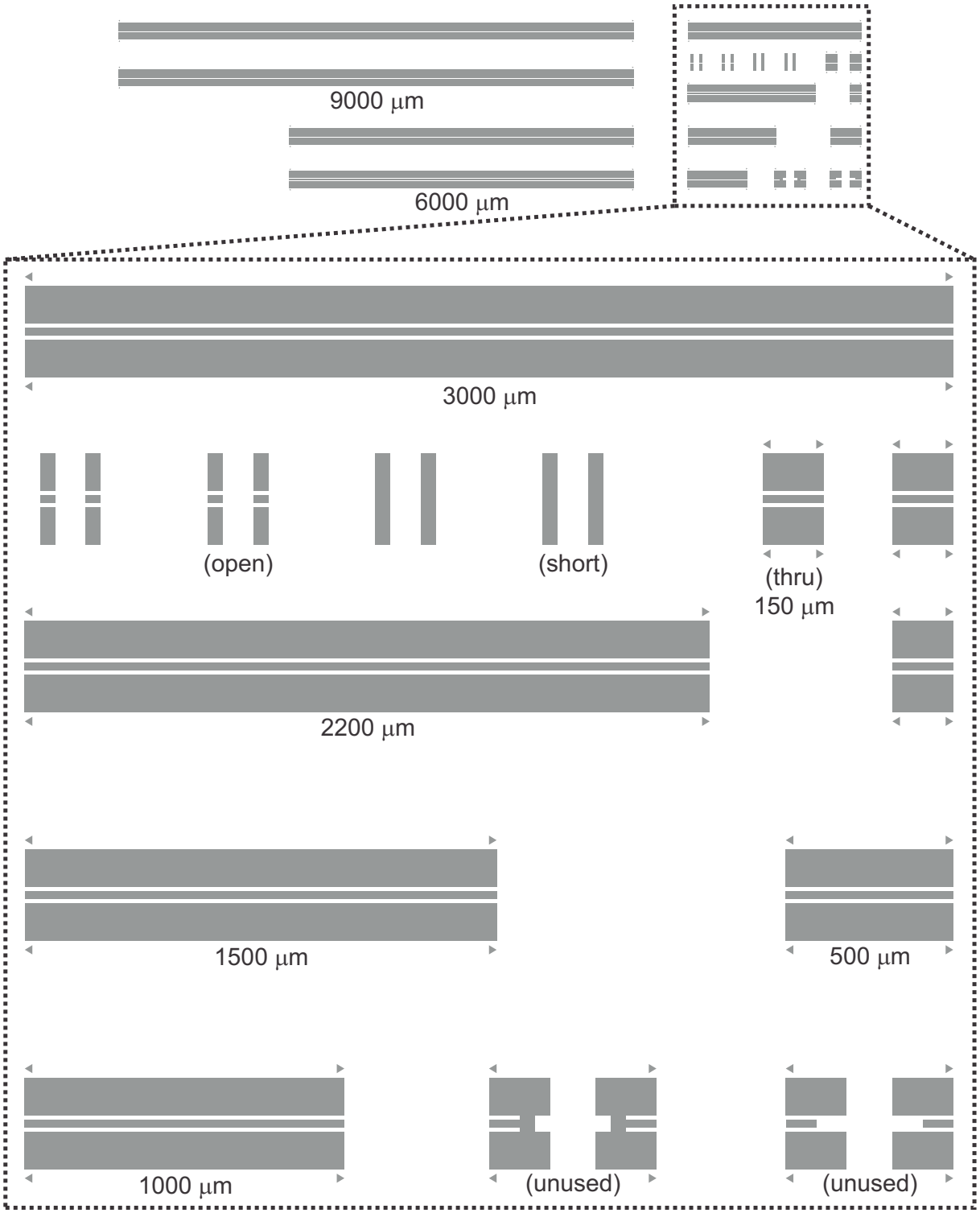

Figure A.1: Schematic of the multiline TRL calibration substrate [11]. CPW dimensions are 24 μm/15 μm/123 μm (widths of signal line/gap/ground line).

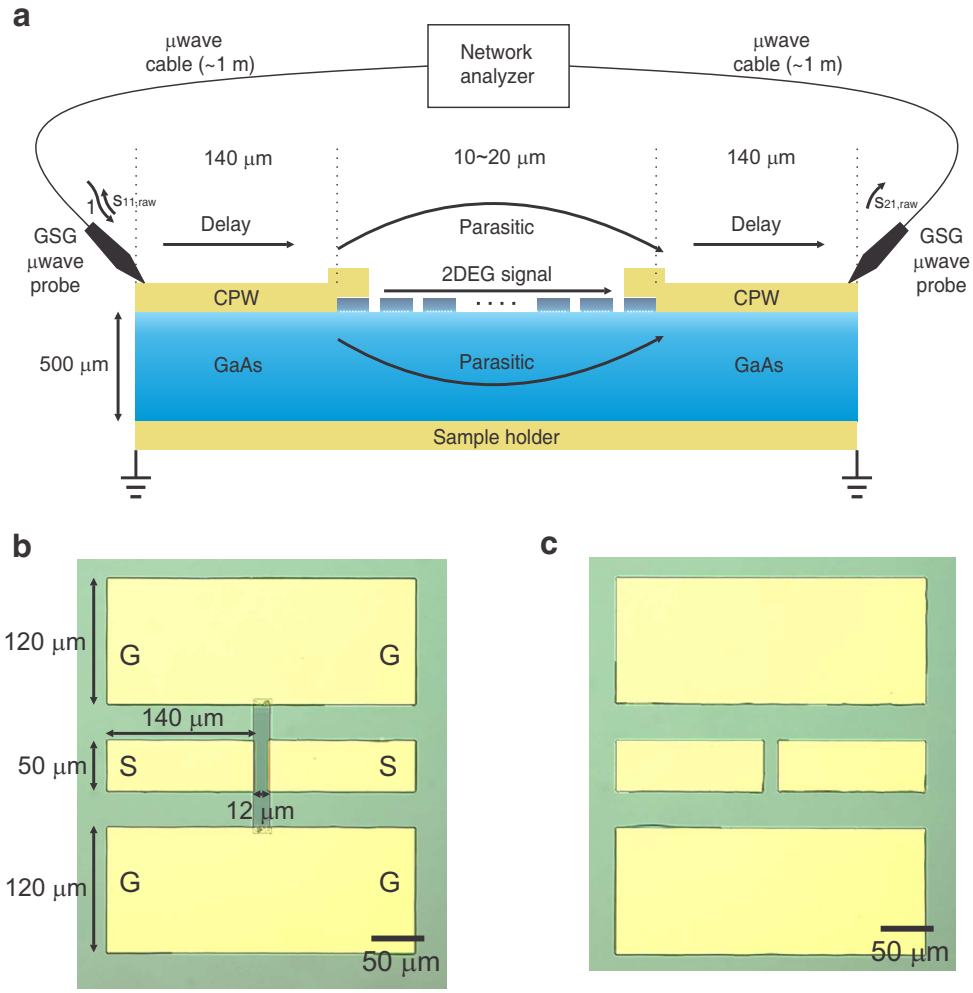

Figure A.2: **a**, Schematic of the microwave measurement setup. The illustration is taken from the 2D plasmonic negative index metamaterial work [14], but it applies similarly to all other devices appearing in the thesis. **b**, Optical image of the device appearing in Fig. 3.12. **c**, The 'open' structure corresponding to the device in **b**.

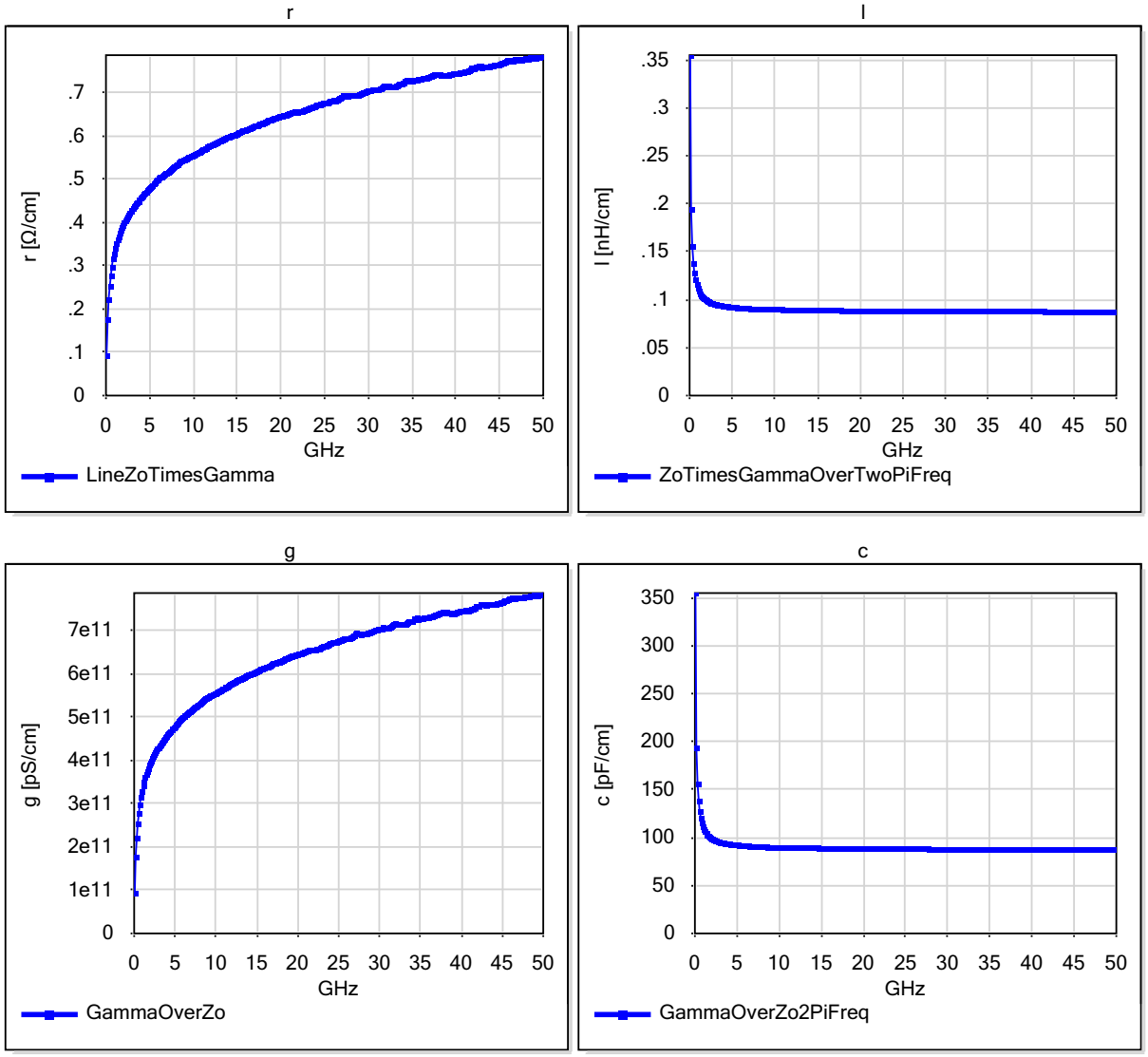

Figure A.3: r-l-g-c parameters of the CPWs extracted from the calibration procedure. Data taken from the 296 K measurements of Ref. [16].

the DUT, and parasitic signals that bypass the DUT. In order to obtain the pure response of the DUT, the delays in the CPWs are first removed from the measured signal by essentially moving the reference planes further to the planes right before the DUT (Fig. A.2a), and then the parasitic signals are de-embedded from the remaining signal by using the parasitic signals measured from an 'open' device (Fig. A.2c). The CPW delays are removed from both the actual device and the 'open' device measurements.

Moving the Reference Planes

Shift of the reference planes is performed by constructing the transmission matrix of an arbitrary length of the CPWs and multiplying the inverse of these matrices to the measured device signals' transmission matrix. The conversion from the measured s-parameters to transmission matrix parameters is performed by [25]

$$\begin{aligned} T_{11} =& \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{2S_{21}}, \\ T_{12} =& Z_0\frac{(1+S_{11})(1+S_{22})-S_{12}S_{21}}{2S_{21}}, \\ T_{21} =& Y_0\frac{(1-S_{11})(1-S_{22})-S_{12}S_{21}}{2S_{21}}, \\ T_{22} =& \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{2S_{21}}, \end{aligned} \tag{A.1}$$

where $Z_0 \equiv 1/Y_0$ is the characteristic impedance of the measurement environment ($Z_0 = 50\ \Omega$ in our case). The transmission matrix for a CPW of length l_i is given by [25]

$$T_{\mathrm{CPW},i} = \begin{bmatrix} \cosh\gamma l_i & z_0 \sinh\gamma l_i \\ y_0 \sinh\gamma l_i & \cosh\gamma l_i \end{bmatrix}, \tag{A.2}$$

where $\gamma = \sqrt{(r+i\omega l)(g+i\omega c)}$ and $z_0 = \sqrt{(r+i\omega l)/(g+i\omega c)} \equiv 1/y_0$. The transmission matrix for the device after removing the delay from length l_1 of the CPW on the left of the device and length l_2 of the CPW on the right (Fig. A.4a) is then given by

$$T_{\mathrm{DUT+parasitic}} = [T_{\mathrm{CPW},1}]^{-1} T_{\mathrm{measured}} [T_{\mathrm{CPW},2}]^{-1}. \tag{A.3}$$

Note that this still includes the effects of the parasitic signals. The same procedure is performed to obtain T_{open}, which is the transmission matrix of the 'open' device (Fig. A.4b) after removing the CPW delay effects. The resulting transmission matrix parameters can be converted back to s-parameters via [25]

$$\begin{aligned} S_{11} =& \frac{T_{11} + T_{12}/Z_0 - T_{21}Z_0 - T_{22}}{T_{11} + T_{12}/Z_0 + T_{21}Z_0 + T_{22}}, \\ S_{12} =& \frac{2(T_{11}T_{22} - T_{12}T_{21})}{T_{11} + T_{12}/Z_0 + T_{21}Z_0 + T_{22}}, \\ S_{21} =& \frac{2}{T_{11} + T_{12}/Z_0 + T_{21}Z_0 + T_{22}}, \\ S_{22} =& \frac{-T_{11} + T_{12}/Z_0 - T_{21}Z_0 + T_{22}}{T_{11} + T_{12}/Z_0 + T_{21}Z_0 + T_{22}}. \end{aligned} \tag{A.4}$$

In the measurements, γl_i is determined at each frequency by examining s_{ii} of the 'open' device (the actual device biased to be nearly open can also be used) instead of using the geometric length, as the actual point of contact of the probe to the CPW is not exactly known. In the measurement of s_{ii} for the 'open' device, the signal travels to the middle and reflects back, traveling a total distance of $2l_i$ with negligible reflection from the probe side as the probes and the CPWs are matched to $50\ \Omega$. Assuming that the termination at the open end is 'open' enough, the impedance looking into the CPW is given by $Z_{\text{in},i} = Z_0/\tanh\gamma l_i$, and hence γl_i can be found by examining

$$s_{ii} = \frac{Z_{\text{in},i} - Z_0}{Z_{\text{in},i} + Z_0} = \frac{1 - \tanh\gamma l_i}{1 + \tanh\gamma l_i}. \tag{A.5}$$

With negligible losses (which is the case with our CPWs on GaAs), this reduces to $\angle s_{ii} = -2\tan\beta l_i$, where $\gamma = i\beta$.

De-embedding the Parasitic Signals

In the microwave measurements, parasitic couplings, such as those originating from the direct couplings between the probes or the CPWs that bypass the DUT, arise and distort the measured signal. In the works appearing in this thesis, we remove this effect by measuring the parasitic signal from the 'open' device and de-embedding it from the measured signal. This is performed by converting the measured s-parameters (with the CPW delays removed) to admittance matrix parameters and subtracting the admittance matrix of the parasitic signals (with the CPW delays removed) from the device's admittance

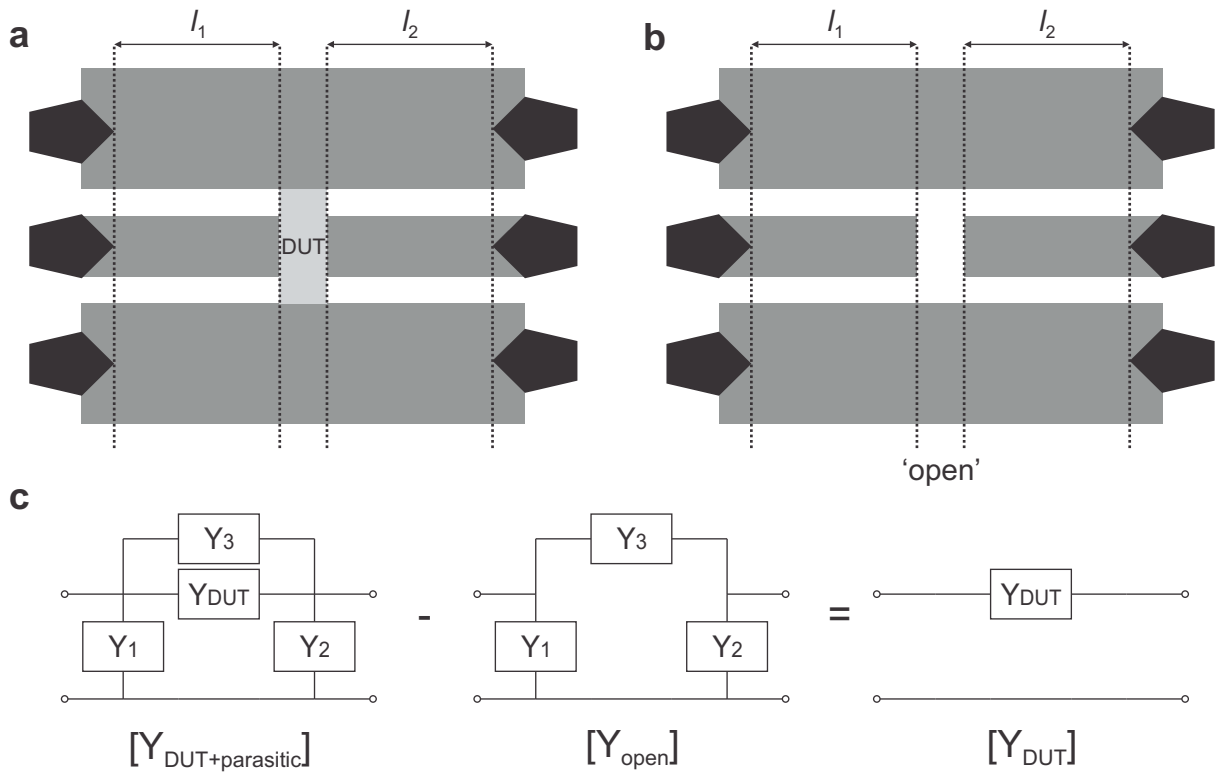

Figure A.4: 'Open' de-embedding scheme. **a**, Illustration of the measured device before de-embedding, with the device-under-test (DUT) at the center. **b**, Illustration of the 'open' structure. **c**, Schematic explanation of the de-embedding procedure.

matrix. The conversion from s-parameters to admittance matrix parameters is performed by [25]

$$\begin{aligned}
Y_{11} &= Y_0 \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}, \\
Y_{12} &= Y_0 \frac{-2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}, \\
Y_{21} &= Y_0 \frac{-2S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}, \\
Y_{22} &= Y_0 \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}.
\end{aligned} \tag{A.6}$$

The de-embedding procedure is illustrated in Fig. A.4c, where Y_1, Y_2, and Y_3 represent the different parasitic signal contributions as circuit components. One can show using simple network analysis theory [25] that the de-embedded admittance matrix of the DUT is given by

$$Y_{\mathrm{DUT}} = Y_{\mathrm{DUT+parasitic}} - Y_{\mathrm{open}}, \tag{A.7}$$

where the effects of Y_1, Y_2, and Y_3 are essentially subtracted, or 'de-embedded,' from the device's response. The resulting admittance matrix parameters can be converted back to s-parameters via [25]

$$\begin{aligned}
S_{11} &= \frac{(Y_0-Y_{11})(Y_0+Y_{22})+Y_{12}Y_{21}}{(Y_0+Y_{11})(Y_0+Y_{22})-Y_{12}Y_{21}}, \\
S_{12} &= \frac{-2Y_{12}Y_0}{(Y_0+Y_{11})(Y_0+Y_{22})-Y_{12}Y_{21}}, \\
S_{21} &= \frac{-2Y_{21}Y_0}{(Y_0+Y_{11})(Y_0+Y_{22})-Y_{12}Y_{21}}, \\
S_{22} &= \frac{(Y_0+Y_{11})(Y_0-Y_{22})+Y_{12}Y_{21}}{(Y_0+Y_{11})(Y_0+Y_{22})-Y_{12}Y_{21}}.
\end{aligned} \tag{A.8}$$

The aforementioned procedure is not perfect and may not completely remove the parasitic signal, because the 'open' device may not be perfectly identical to the actual device minus DUT, and also because of probe landing inconsistencies and temperature fluctuations during measurements. The consequences of the residual parasitic signal was discussed in Sec. 4.1.3.4.

Appendix B

Fabrication Details

Detailed recipes for fabrication steps appearing in the thesis are listed below for future reference. Some portions of the following have appeared in Refs. [11, 14, 16]. William F. Andress, Charles Marcus group (formerly at Harvard University), Philip Kim group (formerly at Columbia University), and James Hone group (Columbia University) are acknowledged for some parts of the following.

Photolithography

1. Sonicate in TCE, acetone, IPA
2. Hotplate $180\ ^{\circ}\mathrm{C}$, 5 min
3. Spin resist (Shipley 18xx series)
   1) 500 rpm, 100 rpm/s, 5 sec
      This step will be omitted in the recipes below, but it is to be understood that this initial acceleration step is there before the main spinning step. Upon entering this first step, apply resist while accelerating, and skip to next step. This leads to more even distribution of resist on small chips.
   2) 4000 rpm, 1000 rpm/s, 40 sec
4. Hotplate $115\ ^{\circ}\mathrm{C}$, 3 min
5. Expose
   1805 $\sim 60\ \mathrm{mJ/cm^2}$
   1818 $\sim 130\ \mathrm{mJ/cm^2}$
6. Develop in CD-30, $\sim 30\ \mathrm{sec}$ (can bake before developing; refer to Shipley datasheet)
7. Stop in DI water

E-beam Lithography

1. Sonicate in TCE, acetone, IPA
2. Hotplate $180\ ^{\circ}\mathrm{C}$, 5 min

PMMA

3. Spin resist (can spin multiple layers depending on need)
   4000 rpm, 1000 rpm/s, 40 sec; hotplate $180\ ^\circ\mathrm{C}$, 10 min
4. Expose
   495K, $\sim 300\ \mu\mathrm{C/cm^2}$ @ $30\ \mathrm{kV}$
   950K, $\sim 450\ \mu\mathrm{C/cm^2}$ @ $30\ \mathrm{kV}$
5. Develop (choose either method)
   DI:IPA=1:3, $\sim 60\ \mathrm{sec}$; stop in IPA (prepare the solution cold)
   MIBK:IPA=1:3, $\sim 90\ \mathrm{sec}$; stop in IPA

PMMA / HSQ etch mask

3. Spin resist
   PMMA 950A2, 4000 rpm, 40 sec; hotplate $180\ ^\circ\mathrm{C}$, 5 min
   HSQ (XR-1541, 2∼6 %), 6000 rpm, 60 sec; hotplate $150\ ^\circ\mathrm{C}$, 2 min
4. Expose $\sim 225\ \mu\mathrm{C/cm^2}$ @ $30\ \mathrm{kV}$
5. Develop in CD-26, $\sim 3\ \mathrm{min}$
6. Etch PMMA in $\mathrm{O_2}$ plasma

GaAs 2DEG Specific Recipes

Gate deposition

1. Perform lithography as detailed above
2. Plasma clean 30 sec, 30 W, 30 sccm $\mathrm{O_2}$
3. Dip in $\mathrm{NH_4OH}$:DI=1:3, 5 sec; rinse in DI
4. Evaporate metal
   Thin gates 5 nm Cr / 30 nm Au
   Thick gates 8 nm Cr / 500 nm Au
   Alignment markers 15 nm Cr / 55 nm Au
5. Liftoff in acetone

Ohmic Contacts

1-3. Same as above
4. Evaporate metal
   5 nm Ni / 20 nm Au / 25 nm Ge / 10 nm Au / 5 nm Ni / 40 nm Au
5. Liftoff in acetone
6. Anneal at $420\ ^\circ\mathrm{C}$, ∼50 sec

Mesa Etching

1. Perform lithography as detailed above
2. Wet etch
   Shipley resist: dip in $\mathrm{H_2SO_4}$:$\mathrm{H_2O_2}$:$\mathrm{H_2O}$=1:8:240; etch rate $4 \sim 5\ \mathrm{nm/sec}$
   PMMA resist: dip in $\mathrm{NH_4OH}$:$\mathrm{H_2O_2}$:$\mathrm{H_2O}$=1:1:150; etch rate $\sim 4\ \mathrm{nm/sec}$
3. Stop in DI water

Bibliography

[1] B. J. F. Lin, D. C. Tsui, M. A. Paalanen, and A. C. Gossard, "Mobility of the two-dimensional electron gas in GaAs-$Al_xGa_{1-x}As$ heterostructures," *Applied Physics Letters*, vol. 45, no. 6, p. 695, 1984.

[2] M. Rohlfing, P. Krger, and J. Pollmann, "Quasiparticle band-structure calculations for C, Si, Ge, GaAs, and SiC using Gaussian-orbital basis sets," *Physical Review B*, vol. 48, pp. 17791–17805, Dec. 1993.

[3] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, "Electronic transport in two-dimensional graphene," *Reviews of Modern Physics*, vol. 83, pp. 407–470, May 2011.

[4] D. Pines and D. Bohm, "A collective description of electron interactions: II. collective vs individual particle aspects of the interactions," *Physical Review*, vol. 85, pp. 338–353, Jan. 1952.

[5] D. Bohm and D. Pines, "A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas," *Physical Review*, vol. 92, pp. 609–625, Nov. 1953.

[6] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface plasmon subwavelength optics," *Nature*, vol. 424, no. 6950, pp. 824–830, 2003.

[7] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," *Nature*, vol. 391, pp. 667–669, Feb. 1998.

[8] A. Polman, "Plasmonics applied," *Science*, vol. 322, pp. 868 –869, Nov. 2008.

[9] D. K. Gramotnev and S. I. Bozhevolnyi, "Plasmonics beyond the diffraction limit," *Nature Photonics*, vol. 4, pp. 83–91, Feb. 2010.

[10] F. Stern, "Polarizability of a two-dimensional electron gas," *Physical Review Letters*, vol. 18, pp. 546–548, Apr. 1967.

[11] W. F. Andress, H. Yoon, K. Y. M. Yeung, L. Qin, K. West, L. Pfeiffer, and D. Ham, "Ultra-subwavelength two-dimensional plasmonic circuits," *Nano Letters*, 2012.

[12] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, and F. Wang, "Graphene plasmonics for tunable terahertz metamaterials," *Nature Nanotechnology*, vol. 6, pp. 630–634, Sept. 2011.

[13] K. Y. M. Yeung, H. Yoon, W. Andress, K. West, L. Pfeiffer, and D. Ham, "Two-path solid-state interferometry using ultra-subwavelength two-dimensional plasmonic waves," *Applied Physics Letters*, vol. 102, p. 021104, Jan. 2013.

[14] H. Yoon, K. Y. M. Yeung, V. Umansky, and D. Ham, "A Newtonian approach to extraordinarily strong negative refraction," *Nature*, vol. 488, pp. 65–69, Aug. 2012.

[15] H. Yoon, K. Y. M. Yeung, P. Kim, and D. Ham, "Plasmonics with two-dimensional conductors," *Philosophical Transactions of the Royal Society A*, vol. 372, p. 20130104, Mar. 2014.

[16] H. Yoon, C. Forsythe, L. Wang, N. Tombros, K. Watanabe, T. Taniguchi, J. Hone, P. Kim, and D. Ham, "Measurement of collective dynamical mass of dirac fermions in graphene," *Nature Nanotechnology*, June 2014 (doi:10.1038/nnano.2014.112).

[17] H. Yoon and D. Ham, "Massive thermal fluctuation of massless graphene electrons," *arXiv:1405.2356 [cond-mat]*, May 2014.

[18] S. Luryi, "Quantum capacitance devices," *Applied Physics Letters*, vol. 52, pp. 501–503, Feb. 1988.

[19] J. Xia, F. Chen, J. Li, and N. Tao, "Measurement of the quantum capacitance of graphene," *Nature Nanotechnology*, vol. 4, no. 8, pp. 505–509, 2009.

[20] P. J. Burke, I. B. Spielman, J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, "High frequency conductivity of the high-mobility two-dimensional electron gas," *Applied Physics Letters*, vol. 76, no. 6, p. 745, 2000.

[21] J. Chauhan and J. Guo, "Assessment of high-frequency performance limits of graphene field-effect transistors," *Nano Research*, vol. 4, pp. 571–579, June 2011.

[22] F. Rana, "Graphene terahertz plasmon oscillators," *Nanotechnology, IEEE Transactions on*, vol. 7, no. 1, pp. 91–99, 2008.

[23] M. Staffaroni, J. Conway, S. Vedantam, J. Tang, and E. Yablonovitch, "Circuit analysis in metal-optics," *Photonics and Nanostructures - Fundamentals and Applications*, vol. 10, pp. 166–176, Jan. 2012.

[24] S. Dröscher, P. Roulleau, F. Molitor, P. Studerus, C. Stampfer, K. Ensslin, and T. Ihn, "Quantum capacitance and density of states of graphene," *Applied Physics Letters*, vol. 96, p. 152104, Apr. 2010.

[25] D. M. Pozar, *Microwave engineering*. J. Wiley, 3 ed., 2005.

[26] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Reviews of Modern Physics*, vol. 81, pp. 109–162, Jan. 2009.

[27] E. H. Hwang and S. Das Sarma, "Dielectric function, screening, and plasmons in two-dimensional graphene," *Physical Review B*, vol. 75, p. 205418, May 2007.

[28] S. Das Sarma and E. H. Hwang, "Collective modes of the massless dirac plasma," *Physical Review Letters*, vol. 102, p. 206412, May 2009.

[29] E. H. Hwang and S. Das Sarma, "Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene," *Physical Review B*, vol. 77, p. 115449, Mar. 2008.

[30] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. New York: Brooks Cole, Jan. 1976.

[31] L. A. Falkovsky, "Optical properties of graphene and IV-VI semiconductors," *Physics-Uspekhi*, vol. 51, p. 887, Sept. 2008.

[32] S. J. Allen, D. C. Tsui, and R. A. Logan, "Observation of the two-dimensional plasmon in silicon inversion layers," *Physical Review Letters*, vol. 38, pp. 980–983, Apr. 1977.

[33] S. J. Allen, H. L. Strmer, and J. C. M. Hwang, "Dimensional resonance of the two-dimensional electron gas in selectively doped GaAs/AlGaAs heterostructures," *Physical Review B*, vol. 28, pp. 4875–4877, Oct. 1983.

[34] I. V. Kukushkin, J. H. Smet, S. A. Mikhailov, D. V. Kulakovskii, K. von Klitzing, and W. Wegscheider, "Observation of retardation effects in the spectrum of two-dimensional plasmons," *Physical Review Letters*, vol. 90, p. 156801, Apr. 2003.

[35] U. Mackens, D. Heitmann, L. Prager, J. P. Kotthaus, and W. Beinvogl, "Minigaps in the plasmon dispersion of a two-dimensional electron gas with spatially modulated charge density," *Physical Review Letters*, vol. 53, pp. 1485–1488, Oct. 1984.

[36] S. Allen Jr., F. Derosa, R. Bhat, G. Dolan, and C. Tu, "Standing charge density waves driven by electron drift in patterned (Al, Ga)As/GaAs heterostructures," *Physica B+C*, vol. 134, pp. 332–336, Nov. 1985.

[37] M. Dyakonov and M. Shur, "Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current," *Physical Review Letters*, vol. 71, pp. 2465–2468, Oct. 1993.

[38] W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, "Terahertz emission by plasma waves in 60 nm gate high electron mobility transistors," *Applied Physics Letters*, vol. 84, pp. 2331–2333, Mar. 2004.

[39] W. Knap, Y. Deng, S. Rumyantsev, and M. S. Shur, "Resonant detection of subterahertz and terahertz radiation by plasma waves in submicron field-effect transistors," *Applied Physics Letters*, vol. 81, pp. 4637–4639, Dec. 2002.

[40] I. Nusinsky and A. A. Hardy, "Band-gap analysis of one-dimensional photonic crystals and conditions for gap closing," *Physical Review B*, vol. 73, p. 125104, Mar. 2006.

[41] C. Sirtori, "Applied physics: Bridge for the terahertz gap," *Nature*, vol. 417, pp. 132–133, May 2002.

[42] J. B. Pendry, "Negative refraction makes a perfect lens," *Physical Review Letters*, vol. 85, pp. 3966–3969, Oct. 2000.

[43] V. G. Veselago, "The electrodynamics of substances with simultaneously negative values of ϵ and μ," *Soviet Physics Uspekhi*, vol. 10, pp. 509–514, Apr. 1968.

[44] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, "Composite medium with simultaneously negative permeability and permittivity," *Physical Review Letters*, vol. 84, pp. 4184–4187, May 2000.

[45] R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction," *Science*, vol. 292, pp. 77 –79, Apr. 2001.

[46] S. Linden, C. Enkrich, G. Dolling, M. W. Klein, J. Zhou, T. Koschny, C. M. Soukoulis, S. Burger, F. Schmidt, and M. Wegener, "Photonic metamaterials: Magnetism at optical frequencies," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, pp. 1097–1105, Dec. 2006.

[47] E. Cubukcu, K. Aydin, E. Ozbay, S. Foteinopoulou, and C. M. Soukoulis, "Electromagnetic waves: Negative refraction by photonic crystals," *Nature*, vol. 423, pp. 604–605, June 2003.

[48] J. Valentine, S. Zhang, T. Zentgraf, E. Ulin-Avila, D. A. Genov, G. Bartal, and X. Zhang, "Three-dimensional optical metamaterial with a negative refractive index," *Nature*, vol. 455, no. 7211, pp. 376–379, 2008.

[49] V. A. Podolskiy and E. E. Narimanov, "Strongly anisotropic waveguide as a nonmagnetic left-handed system," *Physical Review B*, vol. 71, p. 201101, May 2005.

[50] A. J. Hoffman, L. Alekseyev, S. S. Howard, K. J. Franz, D. Wasserman, V. A. Podolskiy, E. E. Narimanov, D. L. Sivco, and C. Gmachl, "Negative refraction in semiconductor metamaterials," *Nat Mater*, vol. 6, pp. 946–950, Dec. 2007.

[51] J. B. Pendry, "A chiral route to negative refraction," *Science*, vol. 306, pp. 1353 –1355, Nov. 2004.

[52] R. Meservey and P. M. Tedrow, "Measurements of the kinetic inductance of superconducting linear structures," *Journal of Applied Physics*, vol. 40, pp. 2028–2034, Apr. 1969.

[53] N. Engheta, "Circuits with light at nanoscales: Optical nanocircuits inspired by metamaterials," *Science*, vol. 317, pp. 1698–1702, 2007.

[54] H. J. Lezec, J. A. Dionne, and H. A. Atwater, "Negative refraction at visible frequencies," *Science*, vol. 316, pp. 430 –432, Apr. 2007.

[55] G. V. Eleftheriades, A. K. Iyer, and P. C. Kremer, "Planar negative refractive index media using periodically L-C loaded transmission lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 2702– 2712, Dec. 2002.

[56] C. Caloz and T. Itoh, "Transmission line approach of left-handed (LH) materials and microstrip implementation of an artificial LH transmission line," *IEEE Transactions on Antennas and Propagation*, vol. 52, pp. 1159– 1166, May 2004.

[57] A. Grbic and G. V. Eleftheriades, "Overcoming the diffraction limit with a planar left-handed transmission-line lens," *Physical Review Letters*, vol. 92, p. 117403, Mar. 2004.

[58] K. J. Binns and P. J. Lawrenson, *Analysis and computation of electric and magnetic field problems*. Oxford,: Pergamon Press, 2d ed., 1973.

[59] X. Chen, T. M. Grzegorczyk, B.-I. Wu, J. Pacheco, and J. A. Kong, "Robust method to retrieve the constitutive effective parameters of metamaterials," *Physical Review E*, vol. 70, p. 016608, July 2004.

[60] D. R. Smith, D. C. Vier, T. Koschny, and C. M. Soukoulis, "Electromagnetic parameter retrieval from inhomogeneous metamaterials," *Physical Review E*, vol. 71, p. 036617, Mar. 2005.

[61] S. P. Burgos, R. d. Waele, A. Polman, and H. A. Atwater, "A single-layer wide-angle negative-index metamaterial at visible frequencies," *Nature Materials*, vol. 9, pp. 407–412, Apr. 2010.

[62] M. Choi, S. H. Lee, Y. Kim, S. B. Kang, J. Shin, M. H. Kwak, K.-Y. Kang, Y.-H. Lee, N. Park, and B. Min, "A terahertz metamaterial with unnaturally high refractive index," *Nature*, vol. 470, pp. 369–373, Feb. 2011.

[63] D. Chanda, K. Shigeta, S. Gupta, T. Cain, A. Carlson, A. Mihi, A. J. Baca, G. R. Bogart, P. Braun, and J. A. Rogers, "Large-area flexible 3D optical negative index metamaterial formed by nanotransfer printing," *Nature Nanotechnology*, vol. 6, pp. 402–407, June 2011.

[64] V. M. Shalaev, "Optical negative-index metamaterials," *Nat Photon*, vol. 1, pp. 41–48, Jan. 2007.

[65] C. M. Soukoulis, S. Linden, and M. Wegener, "Negative refractive index at optical wavelengths," *Science*, vol. 315, pp. 47 –49, Jan. 2007.

[66] Y. M. Meziani, H. Handa, W. Knap, T. Otsuji, E. Sano, V. V. Popov, G. M. Tsymbalov, D. Coquillat, and F. Teppe, "Room temperature terahertz emission from grating coupled two-dimensional plasmons," *Applied Physics Letters*, vol. 92, p. 201108, 2008.

[67] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, and J. Hone, "Boron nitride substrates for high-quality graphene electronics," *Nat Nano*, vol. 5, pp. 722–726, Oct. 2010.

[68] E. A. Shaner and S. A. Lyon, "Time-resolved impulse response of the magnetoplasmon resonance in a two-dimensional electron gas," *Physical Review B*, vol. 66, p. 041402, July 2002.

[69] S. A. Ramakrishna, "Physics of negative refractive index materials," *Reports on Progress in Physics*, vol. 68, pp. 449–521, Feb. 2005.

[70] R. B. Marks, "A multiline method of network analyzer calibration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, pp. 1205– 1215, July 1991.

[71] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov, "Two-dimensional gas of massless Dirac fermions in graphene," *Nature*, vol. 438, pp. 197–200, Nov. 2005.

[72] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, "Experimental observation of the quantum Hall effect and Berry's phase in graphene," *Nature*, vol. 438, pp. 201–204, Nov. 2005.

[73] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nat Mater*, vol. 6, pp. 183–191, Mar. 2007.

[74] F. H. L. Koppens, D. E. Chang, and F. J. Garcia de Abajo, "Graphene plasmonics: A platform for strong light-matter interactions," *Nano Letters*, vol. 11, pp. 3370–3377, Aug. 2011.

[75] J. Chen, M. Badioli, P. Alonso-Gonzlez, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenovi, A. Centeno, A. Pesquera, P. Godignon, A. Zurutuza Elorza, N. Camara, F. J. G. de Abajo, R. Hillenbrand, and F. H. L. Koppens, "Optical nano-imaging of gate-tunable graphene plasmons," *Nature*, vol. advance online publication, June 2012.

[76] A. N. Grigorenko, M. Polini, and K. S. Novoselov, "Graphene plasmonics," *Nature Photonics*, vol. 6, pp. 749–758, Nov. 2012.

[77] H. Yan, X. Li, B. Chandra, G. Tulevski, Y. Wu, M. Freitag, W. Zhu, P. Avouris, and F. Xia, "Tunable infrared plasmonic devices using graphene/insulator stacks," *Nature Nanotechnology*, vol. 7, no. 5, pp. 330–334, 2012.

[78] G. Deligeorgis, M. Dragoman, D. Neculoiu, D. Dragoman, G. Konstantinidis, A. Cismaru, and R. Plana, "Microwave propagation in graphene," *Applied Physics Letters*, vol. 95, pp. 073107–073107–3, Aug. 2009.

[79] H.-J. Lee, E. Kim, J.-G. Yook, and J. Jung, "Intrinsic characteristics of transmission line of graphenes at microwave frequencies," *Applied Physics Letters*, vol. 100, pp. 223102–223102–3, May 2012.

[80] D.-Y. Jeon, K. J. Lee, M. Kim, D. C. Kim, H.-J. Chung, Y.-S. Woo, and S. Seo, "Radio-frequency electrical characteristics of single layer graphene," *Japanese Journal of Applied Physics*, vol. 48, p. 091601, 2009.

[81] L. Wang, I. Meric, P. Y. Huang, Q. Gao, Y. Gao, H. Tran, T. Taniguchi, K. Watanabe, L. M. Campos, D. A. Muller, J. Guo, P. Kim, J. Hone, K. L. Shepard, and C. R. Dean, "One-dimensional electrical contact to a two-dimensional material," *Science*, vol. 342, pp. 614–617, Nov. 2013. PMID: 24179223.

[82] S. H. Abedinpour, G. Vignale, A. Principi, M. Polini, W.-K. Tse, and A. H. MacDonald, "Drude weight, plasmon dispersion, and ac conductivity in doped graphene sheets," *Physical Review B*, vol. 84, p. 045429, July 2011.

[83] C. Jang, S. Adam, J.-H. Chen, E. D. Williams, S. Das Sarma, and M. S. Fuhrer, "Tuning the effective fine structure constant in graphene: Opposing effects of dielectric screening on short- and long-range potential scattering," *Physical Review Letters*, vol. 101, p. 146805, Oct. 2008.

[84] N. Ohba, K. Miwa, N. Nagasako, and A. Fukumoto, "First-principles study on structural, dielectric, and dynamical properties for three BN polytypes," *Physical Review B*, vol. 63, p. 115207, Mar. 2001.

[85] C. Hwang, D. A. Siegel, S.-K. Mo, W. Regan, A. Ismach, Y. Zhang, A. Zettl, and A. Lanzara, "Fermi velocity engineering in graphene by substrate modification," *Scientific Reports*, vol. 2, Aug. 2012.

[86] K. Maex, M. R. Baklanov, D. Shamiryan, F. lacopi, S. H. Brongersma, and Z. S. Yanovitskaya, "Low dielectric constant materials for microelectronics," *Journal of Applied Physics*, vol. 93, pp. 8793–8841, June 2003.

[87] D. Ham and A. Hajimiri, "Concepts and methods in optimization of integrated LC VCOs," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 896–909, June 2001.

[88] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, pp. 97–109, July 1928.

[89] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical Review*, vol. 32, pp. 110–113, July 1928.

[90] R. Kubo, "The fluctuation-dissipation theorem," *Reports on Progress in Physics*, vol. 29, p. 255, Jan. 1966.

[91] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," *Physical Review*, vol. 83, pp. 34–40, July 1951.

[92] A. C. Betz, F. Vialla, D. Brunel, C. Voisin, M. Picher, A. Cavanna, A. Madouri, G. Fève, J.-M. Berroir, B. Plaçais, and E. Pallecchi, "Hot electron cooling by acoustic phonons in graphene," *Physical Review Letters*, vol. 109, p. 056805, Aug. 2012.

[93] A. C. Betz, S. H. Jhang, E. Pallecchi, R. Ferreira, G. Fève, J.-M. Berroir, and B. Plaçais, "Supercollision cooling in undoped graphene," *Nature Physics*, vol. 9, pp. 109–112, Feb. 2013.

[94] R. K. Pathria and P. D. Beale, *Statistical mechanics*. Oxford: Elsevier, 3 ed., 2011.